

THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Diagnostic clinique basé sur l'analyse discriminante logistique

Yapi Marcellin, Adei

Award date:
1996

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Faculté Universitaire Notre Dame de la Paix
Institut d'Informatique

Année académique 95 - 96

ADEI, 4.

**DIAGNOSTIC CLINIQUE BASE
SUR L'ANALYSE
DISCRIMINANTE LOGISTIQUE**

RÉALISATEUR

Adei Yapi Marcellin

COPROMOTEUR

Dr Philippe Beyne

PROMOTEUR

Jean Paul Leclercq

Résumé

Les diagnostics médicaux des patients ont toujours conduit à un problème de classification des patients soit dans un groupe de population malade ou soit dans un groupe de population sain. Pour cette raison la médecine a eu recours à l'analyse discriminante qui est une des méthodes de classification statistique.

L'analyse discriminante est une méthode qui conduit à une règle de classification simple d'individus entre plusieurs populations.

Pour les besoins de la cause nous nous sommes limités à deux populations dont la population des malades et celle des sains.

La méthode discriminante fait l'objet de plusieurs approches dont:

- l'approche classique ou linéaire basée sur des concepts de nature géométrique(distance, axes factoriels discriminants etc...)

Elle consiste à établir une fonction linéaire appelée fonction discriminante sur base d'un échantillon donnée. Cette fonction sera ainsi utilisée pour classer un individu dans un groupe donnée.

- l'approche logistique qui ne tient pas compte de la nature des variables tests, ni des hypothèses spécifiques aux distributions des populations concernées, est basée sur les propriétés bayésiennes. Cette approche consiste à établir d'abord une relation entre le rapport de vraisemblance et la probabilité à posteriori. Puis ensuite à postuler un modèle exponentiel du rapport de vraisemblance dont les composantes épousent une forme linéaire. Et enfin, de retrouver la forme de la fonction logistique, en remplaçant le rapport de vraisemblance par le modèle exponentiel postulé dans la relation établie avec la probabilité à posteriori, et ce après quelques manipulations mathématiques, d'où le nom d'analyse discriminante logistique.

Cette approche logistique de l'analyse discriminante à donner lieu à la construction d'un logiciel sous Windows.

Ce logiciel permettant d'automatiser ce type d'approche, est limité à six variables tests et deux groupes(le groupe des malades et le groupe des sains)

Abstract

The medical diagnostics always conduct at a classification problem of patients either in a group of persons suffering from a given disease or in group of persons not suffering from this disease. For this reason, medicine has decided to use discriminant analysis. The discriminant analysis constitute one of statistical classification's methods.

Discriminant analysis is a method that conducts to simple classification's rule of patients among several groups. For needs of cause, we have decided to limit our subject at two groups of population(disease population and no disease population).

We have several approaches in the discriminant analysis's method:

- classical or linear approach based on the geometry concepts(distance, discriminants factorial axes and so on...).

This approach consists to establish a linear function called discriminant function with a given sample. This function will be used to classify a patient in one group that we have defined.

- logistic approach in which neither variable type nor specific hypothesis of population distribution are considered, is based on bayesian characteristics.

This approach consist at first to establish a relation between likelihood ratio and posterior probability. After this, we postulate a exponential model of likelihood ratio. The components of this exponential model take a linear form.

Finally, we replace likelihood ratio by his exponential model in the relation between likelihood ratio and posterior probability to found Discriminant Logistic Form after some mathematical manipulations.

On based of Discriminant Logistic Analysis, we have constructed a software under Windows. This software, that allows to automate this approach, is limited to six tests and two groups.

Sommaire

	<u>Pages</u>
Avant-propos	9
Introduction	10

PARTIE I THÉORIE & CONTEXTE

Chapitre 1 Les notions de bases de la théorie des probabilités

I	La notion d'événements aléatoires	12
	A Définition	
	B Quelques opérations ensemblistes sur les événements	
II	La notion de probabilité	13
III	Règles de calcul des probabilités et le théorème de Bayes	14
	A Définition	
	B Théorème de Bayes	
	C Exemples d'application	
	D La probabilité de l'indépendance de deux événements	
IV	Notion de variables aléatoires	16
V	Les opérations simples sur les variables aléatoires	18
VI	Quelques exemples de lois de probabilité	19
VII	Problèmes d'échantillonnage et estimations	20

Chapitre 2 Les notions de base de l'algèbre linéaire

I	Les vecteurs	23
	A Définition	
	B Quelques opérations vectorielles	
II	Les matrices	25
	A Définition	
	B Quelques opérations matricielles	

Chapitre 3 Théorie de l'analyse discriminante

Introduction	30
I Tests de laboratoire clinique	31
II Formulation du problème	31
A Cas d'un test unique	
B Cas de tests plusieurs	
III Règles de décision	32
IV Méthodes de décision bayésiennnes en analyse discriminante	36
V Discrimination linéaire	41
A Décomposition de la matrice de dispersions	
B Détermination des axes factoriels discriminants	
C Affectation d'un individu à un groupe	
D Exemple	
VI Discrimination logistique	49
A Règle de classement	
B Détermination de la forme relationnelle entre les probabilités à posteriori et le rapport de vraisemblance	
B-1 Cas d'un test unique	
B-2 Cas de plusieurs tests	
C Détermination de la forme du rapport de vraisemblance	
C-1 Cas d'un test unique	
C-2 Cas de tests multiples	
D Détermination de la forme logistique	
E Estimation des coefficients de la fonction logistique	
E-1 Échantillon composé du mélange de deux groupes	
E-2 Échantillons séparés	
F Exemple d'application	
Conclusion	

PARTIE II MODÉLISATION

Chapitre 1 Modèle de structuration des données

I	Présentation du modèle	73
II	Description des entités et des associations	73
II-1	Les types d'entités	
II-2	Les types d'associations	
II-3	Le schéma entités associations résultant	
III	Contraintes	77
III-1	Contraintes d'intégrités fonctionnelles	
III-2	Contraintes non fonctionnelles	
IV	Scénario de fonctionnement	77

Chapitre 2 Modèle de structuration des traitements

I	Présentation du modèle	79
II	Application à l'analyse clinique	79
II-1	Niveau projet et niveau application	
II-2	Niveau phase et niveau fonction	
	- Phase de création	
	- Phase de suppression	
	- Phase de consultation	
	- Phase de modification	
	- Phase de fonction calcul	

PARTIE III ANALYSE CONCEPTUELLE

Chapitre 1 Présentation générale de l'architecture physique

- | | | |
|----|---------------------------------------|------------------|
| I | Graphe des modules | 87 |
| II | Spécification algorithmique détaillée | (voir à annexes) |

Chapitre 2 Description générale du fonctionnement du logiciel

Item menu Fichier	90
Item menu Informations	97
Item menu Vraisemblance	100
Item menu Allocation	102
Item menu Aide	104

CONCLUSION GENERALE

<i>BIBLIOGRAPHIE</i>	106
-----------------------------	-----

<i>ANNEXES</i>	108
-----------------------	-----

Les notions de base de l'analyse mathématique
Spécifications algorithmique détaillées

AVANT - PROPOS

Ce mémoire de fin d'étude dont le thème est intitulé "Diagnostic clinique basé sur l'analyse discriminante logistique" est l'aboutissement d'un travail de longue haleine mais néanmoins passionnant, qui m'a occupé pendant des mois.

Mon travail a consisté en deux grandes étapes:

- Je devais dans un premier temps cerner toute la théorie de base qui concernait le thème du sujet. Pour cela j'ai eu recours à des documents d'algèbre linéaire, de probabilités et aussi de statistiques.

- Une fois la première étape terminée, je me suis consacré au travail qui m'avait été réellement demandé c'est-à-dire la réalisation d'un logiciel basé sur l'analyse discriminante logistique. Dans cette étape j'ai dû faire appel à mes connaissances acquises au cours de ma formation à l'Institut d'Informatique et à un certain nombre de livres sur la programmation d'ObjectWindows.

Je tiens ici à remercier tous ceux qui de près ou de loin m'ont permis de réaliser ce travail et tout particulièrement mon promoteur monsieur Jean-Paul Leclercq qui m'a permis de mettre au clair certains éléments de la théorie de l'analyse discriminante sans oublier le Dr Beyne de l'hôpital S^t Camille qui m'a fourni les éléments de base de l'étude. Je tiens aussi à remercier mes compatriotes ivoiriens notamment Achy Mené Gabriel et Niamba Narcisse qui m'ont moralement soutenu tout au long de ce travail, sans oublier les membres de ma famille.

Je tiens enfin à remercier tout le corps enseignant de l'Institut d'Informatique de l'Université Notre Dame de la Paix pour la formation qui m'a été donnée.

Adei y. m.

INTRODUCTION

L'évolution de la médecine et la complexité des maladies ont donné lieu à la mise en place des tests médicaux de plus en plus nombreux pour une maladie donnée.

Ainsi, face aux exigences de l'environnement qui demandent des diagnostics de plus en plus précis et la présence de tests cliniques de plus en plus nombreux donnant lieu souvent à des résultats complémentaires ou contradictoires, le médecin se voit entraîner dans un travail de plus en plus difficile pour décider de l'état de son patient.

C'est dans ce but que la médecine a fait recours aux techniques statistiques et plus particulièrement aux techniques d'analyse discriminante pour les aider dans leur tâche de décision difficile et aléatoire.

Il faut remarquer que même si ces techniques permettent de donner un coup de main aux médecins, force est de remarquer que ces techniques fournissent eux aussi des résultats aléatoires. Ces résultats sont quelques fois incomplets car ils n'intègrent pas tous les paramètres possibles sur le patient comme par exemple les paramètres que l'on ne peut pas modéliser.

L'analyse discriminante qui intéresse la médecine est une technique de classification des objets basée sur des règles de décision.

Cette analyse discriminante fait l'objet de plusieurs approches dont l'approche classique ou linéaire, l'approche logistique etc...

L'approche qui nous intéresse plus ici est l'approche logistique qui est basée sur l'établissement d'une relation directe entre la probabilité *a posteriori* et le rapport de vraisemblance.

Avant d'entrée dans le but du sujet, je tiens à signaler que le fascicule est divisé en trois grandes parties dont la première partie concerne la théorie de l'analyse discriminante et les deux dernières touchent la programmation de l'analyse logistique discriminante.

Cette première partie est composée de trois chapitres dont les deux premiers portent sur des rappels brefs de la théorie des probabilités et de l'algèbre linéaire, et le troisième chapitre concernant la théorie de l'analyse discriminante.

La deuxième partie du fascicule quant elle se compose de deux chapitres dont l'un concerne le modèle de structuration des données et l'autre le modèle de structuration des traitements.

La troisième partie comporte lui aussi deux chapitres dont les spécifications formelles et la présentation générale de l'architecture physique.

PARTIE I

THEORIE & CONTEXTE

Chapitre I

LES NOTIONS DE BASE DE LA THEORIE DES PROBABILITES

I - La notion d'événements aléatoires.

A) Définition:

Le concept d'événements aléatoire inclut l'idée d'une expérience c'est-à-dire une procédure dont les résultats ne peuvent être prédire avec certitude. Les exemples les plus familiers en statistique sont le lancement d'une pièce équilibrée, le lancé de dé(s) etc... ,qui sont considérés comme des expériences.

Le fait qu'une expérience répétée dans des conditions apparemment identiques est susceptible de nous fournir des résultats différents, sans que nous puissions les prévoir nous pousse à dire que l'expérience et ses résultats sont aléatoires.

Chaque expérience réalisée produit un résultat élémentaire appelé événement élémentaire e_j .

L'ensemble des événements élémentaires est appelé ensemble fondamental μ .

Un sous ensemble de μ est appelé événement complexe ou tout simplement événement.

Les événements susceptibles de se produire sont appelés événements possibles, dans le cas contraire ils sont qualifiés événements impossibles.

Lorsqu'un événement se produit effectivement on le qualifie événement réalisé.

Ainsi donc:

e_i : sera qualifié de $i^{\text{ème}}$ événement élémentaire.

$\mu = \{e_1, e_2, \dots, e_i, \dots, e_n\}$ est qualifié d'ensemble fondamental ou ensemble de tous les événements élémentaires ou encore événement certain c'est-à-dire un événement qui se réalise toujours .

$S = \{e_1, e_2, e_5\}$, $T = \{e_1, e_7, e_8, e_{10}\}$, $W = \{e_3, e_{11}\}$ etc... sont des sous ensembles de μ donc des événements de μ .

\emptyset = est appelé ensemble vide ou événement impossible.

$\text{non } S = \neg S = \{e_3, e_4, e_6, e_7, \dots, e_i, \dots, e_n\}$ est qualifié de complément de S.

B) Quelques opérations ensemblistes sur les événements

$S \text{ ou } T = (S \cup T) = \{e_1, e_2, e_5, e_7, e_8, e_{10}\}$ est appelé l'union de S et T.

$S \text{ et } T = (S \cap T) = \{e_1\}$ est appelé l'intersection de S et de T.

$S \text{ et } W = (S \cap W) = \emptyset$ S et W sont incompatibles.

$S \text{ et } \neg S = (S \cap \neg S) = \emptyset$ S et $\neg S$ sont incompatibles.

En prenant l'exemple du lancé de deux dés où chaque dé à six faces.

On peut définir ainsi plusieurs objets par exemple:

- faire la somme des nombres sur chaque face des dés
- prendre le plus grand nombre sur chaque face des dés
- faire la différence des nombres sur chaque face des dés
- etc...

Les objets ainsi définis sont appelés variables aléatoires.

On associe à toute variable aléatoire un ensemble de résultats possibles ou réalisables lors de l'expérience.

Dans le cas de notre exemple de dés , en choisissant l'objet « faire la somme des nombres sur chaque face des dés » alors cette variable aléatoire sera associée à onze résultats entiers possibles allant de 2 à 12.

- ensemble fondamentale $\mu = \{2, 3, 4, 5, 6, \dots, 12\}$
- événements élémentaires 2 ou 3 ou 4 etc...
- événements $T = \{2\}$ $S = \{2, 4\}$ $N = \{7, 9, 11\}$ etc...

II - La notion de probabilité

Lorsqu'on répète N fois une expérience et qu'un résultat apparaît n fois, le rapport n/N définit la fréquence d'apparition de ce résultat. Cette fréquence est toujours comprise entre 0 et 1. D'une manière générale , lorsque le nombre des épreuves augmentent indéfiniment, on constate que la fréquence d'un résultat tend vers une limite qui apparaît comme une caractéristique de l'épreuve.

L'existence d'une telle limite à la fréquence relative d'un événement aléatoire suggère que cet événement présente en lui même , dans les condition d'une épreuve donnée, des caractéristiques constantes que l'on peut exprimer par un nombre, compris entre 0 et 1.

Ce nombre sera qualifié de probabilité de l'événement .

La probabilité de réalisation d'un événement quelconque S au cours d'une expérience sera noté $P(S)$. Cette probabilité mesure le nombre de chances que l'événement S puisse se réaliser cours d'une expérience donnée.

La probabilité peut être définie comme une valeur réelle positive qui vérifie les caractéristiques suivantes :

$$P(\emptyset) = 0 \text{ où } \emptyset \text{ est un événement impossible.}$$

$$\forall T \in \mu \quad 0 \leq P(T) \leq 1$$

$$P(\mu) = 1 \text{ où } \mu \text{ est considéré comme un événement certain}$$

$$\forall T \in \mu \text{ et } \forall S \in \mu \text{ tel que } T \cap S = \emptyset \text{ on a alors } P(T \cup S) = P(T) + P(S)$$

En revenant dans l'expérience des lancés de dés avec objet '' faire la somme des nombres sur chaque face des dés'' nous permet de déterminer ainsi la liste des résultats (événements élémentaires) et les probabilités qui leur sont associées dans le tableau ci-dessous:

RESULTATS	PROBABILITES	PROBABILITES CUMULEES
2	1/36	1/36
3	2/36	3/36
4	3/36	6/36
5	5/36	10/36
6	5/36	15/36
7	6/36	21/36
8	5/36	26/36
9	4/36	30/36
10	3/36	33/36
11	2/36	35/36
12	1/36	36/36

III - Règles de calcul des probabilités et théorème de Bayes

A) Règles de calcul

- La probabilité de la réunion de deux événements S et T est égale à la somme de leurs probabilités diminuée de la probabilité de leur intersection

$$P(S \cup T) = P(S) + P(T) - P(S \cap T).$$

$P(S \cup T)$ est qualifiée de probabilité totale de S et T alors que $P(S \cap T)$ est qualifiée de probabilité composée.

Dans l'éventualité où l'événement S est réalisé, la probabilité de réalisation de l'événement T est qualifiée de probabilité conditionnelle de T sachant S réalisée est notée $P(T|S)$ et vis-versa.

Si les événements S et T sont incompatibles alors $P(S \cap T) = \emptyset$ cela entraîne le fait que

$$P(S \cup T) = P(S) + P(T).$$

Ainsi $P(S \cap T)$ peut s'exprimer comme suit:

$$P(S \cap T) = P(S).P(T|S) = P(T).P(S|T)$$

D'où on peut déduire la formule de la probabilité conditionnelle de l'événement S sachant l'événement l'événement T déjà réalisé

$$P(S|T) = \frac{P(S \cap T)}{P(T)}$$

B) Théorème de Bayes

Soit un événement A susceptible de se réaliser selon n modalités différentes et incompatibles $A_1, A_2, \dots, A_i, \dots, A_n$ et un événement B qui n'existe qu'en intersection avec A.

Chaque modalité A_i de l'événement A est en intersection avec l'événement B soit $A_i \cap B$.

D'après l'explication précédente la probabilité de $A_i \cap B$ serait égale à

$$P(A_i \cap B) = P(A_i) \cdot P(B|A_i).$$

Par hypothèse B qui est l'union des $(A_i \cap B)$ «ces termes sont incompatibles» donc

$$P(B) = P(\sum (A_i \cap B)) = \sum P(A_i \cap B) = \sum P(A_i) \cdot P(B|A_i).$$

En considérant un résultat final $A_k \cap B$ on aura

$$P(A_k|B) = P(A_k) \cdot \frac{P(B|A_k)}{P(B)} \quad (1)$$

En remplaçant P(B) par sa valeur dans (1) on obtient

$$P(A_k|B) = P(A_k) \cdot \frac{P(B|A_k)}{\sum_i P(A_i) \cdot P(B|A_i)}$$

Ce qui constitue le théorème de Bayes.

L'intérêt de ce théorème est de mettre, en présence du résultat $(A \cap B)$ de deux événements dont l'un est bien défini (B) mais l'autre a pu se réaliser selon diverses modalités, de calculer la probabilité que ce soit une modalité donnée (A_k) qui soit en cause.

Cela exige que l'on connaisse les probabilités de ces modalités $[P(A_i)]$, qualifiées de probabilités à priori (on entend par là sans considération de B), ainsi que les probabilités des résultats finaux $[P(B|A_i)]$ qualifiées de probabilités à posteriori.

C) Un exemple d'application:

Prenons l'exemple d'un médecin, qui a décidé d'examiner 700 élèves d'un lycée, dont les élèves sont à 90% en bonne santé, mais il a dû décider d'examiner aussi 300 autres élèves d'un second lycée, dont les élèves ne sont en bonne santé que dans une proportion de 80%.

Quelle est la probabilité qu'un élève qui n'est pas en bonne santé vienne du deuxième lycée?

Les événements en présence sont:

A_1 élève en bonne santé du 1^{er} lycée;

A_2 élève en bonne santé du 2^{er} lycée;

B élève malade;

On connaît

$$P(A_1) = 700/1000 = 0.70;$$

$$P(A_2) = 300/1000 = 0.30;$$

$$P(B|A_1) = 0.10;$$

$$P(B|A_2) = 0.20;$$

La probabilité cherchée est :

$$P(A_2|B) = 0.30 \cdot 0.20 / (0.70 \cdot 0.10) + (0.30 \cdot 0.20) = 0.06 / 0.13 = 46\%$$

D) La probabilité de l'indépendance de deux événements.

Deux événements A et B sont qualifiés d'indépendants si les conditions de réalisation de chaque événement est indépendant des conditions de réalisation de l'autre événement.

En d'autre terme $P(A|B) = P(A)$ et $P(B|A) = P(B)$.

Les conséquences de cette définition sont que

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) = P(A) \cdot P(B).$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A) \cdot P(B).$$

IV - Notion de variables aléatoires

Précédemment, nous avons considéré un événement aléatoire auquel nous avons associé une probabilité. Et nous avons dit par « événement » qu'il fallait entendre non seulement l'apparition d'un phénomène, mais la prise en considération d'une certaine valeur numérique x par une variable X . C'est sur ce dernier cas que nous allons insister.

Par définition nous appellerons variable aléatoire X une variable susceptible de prendre différentes valeurs x auxquelles il est possible d'affecter une probabilité.

On peut distinguer deux grands types de variables aléatoires dont les variables aléatoires dites discrètes et celles qualifiées de continues.

Cas des variables aléatoires discrètes:

La variable aléatoire discrète est une variable susceptible de prendre des valeurs bien définies et fixes.

La fonction qui permet de calculer la probabilité d'une valeur prise par la variable aléatoire est appelé fonction de masse ou loi de probabilité notée

$$P(x) = P(X=x) \quad \text{avec } 0 \leq P(x) \leq 1$$

et

$$\sum_x P(x) = 1.$$

La fonction F qui permet de déterminer une valeur de la variable aléatoire X inférieure ou égale à b est qualifiée de fonction de répartition ou de distribution de la variable

$$F(b) = P(X \leq b) = \sum_{x \leq b} P(x) \quad \text{fonction de répartition non décroissante}$$

avec $F(-\infty) = 0$ et $F(+\infty) = 1$.

$$P(a \leq X \leq b) = F(b) - F(a-1).$$

On appelle moyenne ou espérance mathématique d'une variable aléatoire discrète l'expression suivante

$$E(X) \equiv \mu = \sum_x P(x)x.$$

On appelle variance d'une variable aléatoire discrète l'expression suivante

$$V(X) \equiv \sigma^2 = E[(X - \mu)^2] = \sum_x P(x)(x - \mu)^2 = \sum_x P(x)x^2 - \mu^2$$

Cette expression est toujours positive ou nulle et la racine carrée celle-ci est appelée écart type mesurant la moyenne des écarts des valeurs prises par cette variable aléatoire X par rapport à l'espérance mathématique.

Cas des variables aléatoires continues:

La variable aléatoire continue est une variable susceptible de prendre des valeurs bien définies et continues.

La fonction qui permet de calculer la probabilités d'une valeur prise par la variable aléatoire est appelée fonction de masse ou loi de probabilité et notée:

$d(x)$ la fonction ou loi de densité de probabilité avec $0 \leq d(x)$

et

$$\int d(x)dx = 1.$$

La fonction F qui permet de déterminer une valeur de la variable aléatoire X inférieure ou égale à b est qualifiée de fonction de répartition ou de distribution de la variable

$$F(b) = P(X \leq b) = \int^b d(x)dx \quad \text{avec } F(-\infty) = 0 \text{ et } F(+\infty) = 1$$

$$F(b) - F(a) = P(a \leq X \leq b)$$

On appelle moyenne ou espérance mathématique d'une variable aléatoire continue l'expression suivante

$$E(X) \equiv \mu = \int x d(x) dx.$$

On appelle variance d'une variable aléatoire continue l'expression suivante

$$V(X) \equiv \sigma^2 = E[(X - \mu)^2] = \int (x - \mu)^2 d(x) dx = \int x^2 d(x) dx - \mu^2$$

Cette expression est toujours positive ou nulle et la racine carrée celle-ci est appelée écart type mesurant la moyenne des écarts des valeurs prises par cette variable aléatoire X par rapport à l'espérance mathématique.

V - Les opérations simples sur les variables aléatoires

L'espérance mathématique et la variance d'une fonction linéaire de variable aléatoire:

Soient une variable aléatoire X et un coefficient a , le produit de a et X donne une nouvelle variable aléatoire soit $Z = a.X$.

L'espérance mathématique de cette nouvelle variable aléatoire Z donne

$$E(Z) = E(a.X) = a.E(X) = a.\mu_X$$

La variance de cette nouvelle variable aléatoire Z donne

$$V(Z) = V(a.X) = a^2.V(X) = a^2.\sigma_X^2$$

L'espérance mathématique et la variance d'une somme de deux variables aléatoires:

Soient deux variables aléatoires X et Y , indépendantes leur somme donne une nouvelle variables aléatoires $Z = X + Y$ dont l'espérance mathématique est égale à la somme des espérances mathématiques des termes

$$\mu_Z = E(Z) = E(X + Y) = E(X) + E(Y) = \mu_X + \mu_Y \text{ (valable même quand } X \text{ et } Y \text{ ne sont pas indépendants)}$$

La variance de cette nouvelle variable aléatoire Z donne

$$\sigma_Z^2 = V(Z) = V(X + Y) = V(X) + V(Y) = \sigma_X^2 + \sigma_Y^2$$

VI - Quelques exemples de lois de probabilités

1) Dans le cas de variables aléatoires discrètes on peut prendre comme exemple la loi binomiale.

La distribution binomiale est une distribution discontinue où la variable ne peut prendre que des valeurs entières, de 0 à $+\infty$, dont les probabilités sont données par les termes successifs du développement de $(q + p)^n$.

Ainsi la probabilité d'avoir la valeur x_i est donnée par

$$P(X = x_i) = P(x_i) = \frac{n!}{i!(n-i)!} q^{n-i} p^i$$

Espérance mathématique de x_i : $\mu = E(X) = np$.

Variance x_i : $\sigma^2 = npq$

2) Dans le cas de variables aléatoires continues on peut prendre comme exemple la loi normale. La distribution normale comme toutes les distributions continues, est définie par la densité de probabilité.

Cette fonction de densité de probabilité définie par

$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} \text{Exp}\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

avec $-\infty < x < +\infty$

est la loi de densité de probabilité les plus célèbres.

L'espérance mathématique de cette loi de densité est

$$E(X) \equiv \mu \text{ avec } -\infty < \mu < +\infty$$

La variance de cette loi de densité est

$$V(X) \equiv \sigma^2 \text{ avec } 0 \leq \sigma < +\infty$$

En opérant un changement de variable de telle sorte que la moyenne soit nulle ($\mu = 0$) c'est-à-dire en posant $x - \mu$ et que la variance soit égale à 1

($\sigma = 1$) c'est-à-dire en posant $Z = \left(\frac{x-\mu}{\sigma}\right)$ nous obtenons une loi de probabilité plus simple appelée la loi normale réduite (ou Laplace-Gauss) comme suite

$$P(Z) = \frac{1}{\sqrt{2\pi}} \text{Exp}\left[-\frac{1}{2}Z^2\right]$$

VII - Problèmes d'échantillonnage et estimations.

La population est l'ensemble de tous les individus qui relèvent d'une définition donnée, le terme individu étant pris au sens large et pouvant désigner un objet, un phénomène, un nombre etc... La population est plus ou moins vaste selon la définition qu'on adopte pour les individus qui la composent.

La population est donc l'ensemble des données que l'on considère, dont l'effectif est plus ou moins grand suivant le point de vue où l'on se place, mais qui bien souvent est infini.

L'échantillon, au contraire, est la fraction de la population qui a fait l'objet des observations, le prélèvement étant fait au hasard.

L'effectif de l'échantillon est souvent nommé la taille de l'échantillon.

La population d'où est tirée l'échantillon est appelée population parente.

Il faut souligner que l'extraction d'un échantillon de cette population doit être fait absolument au hasard.

Estimation de la moyenne

Soit une population (d'effectif infini) et un échantillon de taille n .

Soit μ la moyenne de la population et μ' la moyenne de l'échantillon, cette dernière étant dite moyenne empirique (c'est-à-dire fournie par l'observation ou l'expérience).

On démontre que l'on peut prendre pour estimation de la moyenne μ , ou encore moyenne estimée, la moyenne empirique elle-même. On a donc:

$$m = \mu' = \frac{\sum_i X_i}{n} = \frac{\sum_i f_i X_i}{n}$$

Cette estimation est absolument correcte car:

$$E(\mu') = \frac{\sum_i E(X_i)}{n} = \frac{n\mu}{n} = \mu$$

$$(\sigma_{\mu'})^2 = \frac{\sum_i (\sigma_{X_i})^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \text{ avec } (\sigma_{\mu'})^2 \rightarrow 0 \text{ quand } n \rightarrow +\infty$$

Estimation de la variance

Pour estimer σ^2 , variance de la population, il est préférable de prendre non pas σ'^2 , variance de l'échantillon ou variance empirique qui est égale à:

$$\sigma'^2 = \frac{\sum_i (X_i - \mu')^2}{n}$$

et dont l'espérance mathématique

$$E(\sigma'^2) = \frac{(n-1)\sigma^2}{n} \neq \sigma^2 \quad \text{quelque soit } n$$

mais par contre en prenant

$$S^2 = \frac{n\sigma'^2}{n-1} = \frac{\sum_i (X_i - \mu')^2}{n-1}$$

on a $E(S^2) = \sigma^2$ donc S^2 représente l'estimateur absolument correct de σ^2 , quelque soit n .
Cependant, comme en général la moyenne μ de la population reste inconnu alors on la remplace dans l'expression de S^2 par son estimateur m ce qui nous donne

$$S'^2 = \frac{\sum_i (X_i - m)^2}{n-1}$$

Quelques exemples d'applications

1) on considère une population n dans laquelle la probabilité qu'un individu pris quelconque soit malade est identique pour tous les patient et vaut p . La probabilité que ce dernier ne soit pas malade est de $(1-p)$.

Quelle est la probabilité d'avoir dans cette population n_1 malades et n_2 non malades avec $(n_1 + n_2 = n)$?

Cette probabilité est $P(n_1 \in D, n_2 \in \bar{D}) = p^{n_1} \cdot (1-p)^{n_2}$.

Ce modèle est appelé fonction de vraisemblance des résultats observés et est noté

$$L(p) = p^{n_1} \cdot (1-p)^{n_2}.$$

Le principe du maximum de vraisemblance est de trouver ou d'estimer la valeur de p pour laquelle étant donnée le nombre d'observations n_1 et n_2 , on arrive maximiser le membre de droite de l'équation de vraisemblance ci-dessus.

Dans la pratique il est plus simple et équivalent mathématiquement de trouver la valeur de p qui maximise le logarithme de la vraisemblance:

$$\text{Log}(L(p)) = n_1 \cdot \text{Log}(p) + n_2 \cdot \text{Log}(1-p)$$

$$\frac{\partial \log(L(p))}{\partial p} = n_1 \frac{\partial \text{Log}(p)}{\partial p} + n_2 \frac{\partial \text{Log}(1-p)}{\partial p} = 0$$

$$\Rightarrow p^* = \frac{n_1}{n_1 + n_2} \quad \text{est l'estimateur du maximum de vraisemblance de } p.$$

2) Considérons un autre exemple où on réalise des mesures de sodium par patient sur ensemble de n patients.

Ces mesures représentent un échantillon aléatoire d'une population normale distribuée de mesures.

La fonction de vraisemblance indépendante de ces observations est:

$$L(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i)$$

qui est le produit des fonctions de densités de probabilités individuels.

La fonction de vraisemblance selon l'hypothèse spécifique est

$$L(\mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} (x - \mu)^2\right]$$

Pour tout échantillon de taille n , l'expression ci-dessus est seulement une fonction de paramètre μ et σ^2 .

L'estimateur du maximum de vraisemblance de σ^2 peut être trouver par différentiel de

$$\log(L(\mu, \sigma^2)) = \left(-\frac{n}{2}\right) (\log(2\pi) + \log(\sigma^2)) - (2\sigma^2)^{-1} \sum_{i=1}^n (x_i - \mu)^2$$

par rapport à σ^2

et en annulant cette dérivé

$$\frac{\partial \log(L(\mu, \sigma^2))}{\partial \sigma^2} = 0$$

on obtient

$$\sigma^{*2} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

En général μ n'est pas connu, il est estimé par la moyenne de l'échantillon \bar{x} . Pour compenser le degré de liberté le dénominateur n est réduit de 1 et l'estimateur recherché donne

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Cet estimateur est utilisé à la place de σ^{*2} .

Ainsi donc la variance de l'échantillon S^2 n'est pas un estimateur du maximum de vraisemblance du paramètre σ^2 dans la distribution normale mais la meilleur approximation disponible pour ça.

Chapitre II

LES NOTIONS DE BASE DE L'ALGEBRE LINEAIRE

I - Les vecteurs

A) Définition

Un profil de laboratoire est une collection de valeurs de test X_1, X_2, \dots, X_k traité ici comme des variables aléatoires.

En notation mathématique, un tel profil définit un vecteur $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}$

Un vecteur peut être représenté par des lettres majuscules caractérisant des variables aléatoires ou des lettres minuscules caractérisant les valeurs prises par ces variables aléatoires. Ainsi par définition un vecteur est toujours une « colonne d'éléments » et X_1, X_2, \dots, X_k ou x_1, x_2, \dots, x_k sont les composants de ce vecteur.

On appelle transposé d'un vecteur X , le vecteur X^T dont les composants sont en ligne:

$$X^T = (X_1, X_2, \dots, X_k)$$

Le nombre k définit la dimension du vecteur.

Exemple: Soit le vecteur $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ signifiant un profil de trois tests où

X_1 = urée;

X_2 = acide urique

X_3 = créatine

Ce vecteur X représente un ensemble de variables de variables aléatoires.

Un patient avec 5.4 mmol/l, 298 d'acide urique et 78 μ mol/l de créatine, est caractérisé par le

vecteur $x = \begin{bmatrix} 5.4 \\ 298 \\ 78 \end{bmatrix}$

Ce vecteur x représente un ensemble de valeurs prises par les variables définies ci-dessus.

$$x^T = (5.4, 298, 78)$$

B) Quelques opérations vectorielles

Addition et soustraction de vecteurs

On peut additionner ou soustraire deux ou plusieurs vecteurs entre eux qu'à condition qu'ils aient tous la même dimension.

$$\begin{aligned}X^T + Y^T &= (X_1, X_2, \dots, X_k) + (Y_1, Y_2, \dots, Y_k) \\&= (X_1 + Y_1, X_2 + Y_2, \dots, X_k + Y_k) \\&= (X + Y)^T\end{aligned}$$

$$\begin{aligned}X^T - Y^T &= (X_1, X_2, \dots, X_k) - (Y_1, Y_2, \dots, Y_k) \\&= (X_1 - Y_1, X_2 - Y_2, \dots, X_k - Y_k) \\&= (X - Y)^T\end{aligned}$$

La multiplication de vecteurs par un scalaire

On peut multiplier un ou plusieurs vecteurs par scalaire.

$$\gamma X^T = \gamma (X_1, X_2, \dots, X_k) = (\gamma X_1, \gamma X_2, \dots, \gamma X_k) = (\gamma X)^T$$

La multiplication de vecteurs

On peut multiplier deux ou plusieurs vecteurs entre eux qu'à condition qu'ils aient tous la même dimension.

$$\begin{aligned}X^T \cdot Y &= (X_1, X_2, \dots, X_k) \cdot \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{pmatrix} = (X_1 \cdot Y_1, X_2 \cdot Y_2, \dots, X_k \cdot Y_k) \\&= \sum_{i=1}^k X_i Y_i\end{aligned}$$

La moyenne des composants vectorielles

Si les composants du vecteur x représentent des notes de n individus, alors on peut calculer une moyenne de n individus appelée moyenne vectorielle

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

II - Les matrices

A) Définition

une matrice A est un tableau rectangulaire de valeurs appelées composants

de la matrice constituée de r ligne et c colonnes: $A = \begin{vmatrix} a_{11} & . & . & . & a_{1c} \\ . & . & . & . & . \\ . & . & a_{ij} & . & . \\ . & . & . & . & . \\ a_{r1} & . & . & . & a_{rc} \end{vmatrix}$

Quand on a des observations d'un vecteur $X^T = (X_1, X_2, \dots, X_c)$ de c variables aléatoires disponibles pour r sujets; cela forme une matrice de r lignes (sujets) et de c colonnes (variables). Ainsi, on peut affirmer qu'un vecteur est une matrice de r lignes et d'une colonne.

La dimension de la matrice est égale au nombre de lignes multiplié par le nombre de colonne c'est-à-dire $r \times c$.

Ainsi donc toute matrice dont le nombre de lignes est égal au nombre de colonnes est appelé matrice carré.

Le $(i,j)^{\text{ème}}$ élément d'une matrice A est l'élément d'intersection de la $i^{\text{ème}}$ ligne et de la $j^{\text{ème}}$ colonne de cette même matrice A . On représente simplement cet élément par $(A)_{ij}$ ou a_{ij} .

B) Quelques opérations matricielles

Soit deux matrices A et B de même dimension c'est-à-dire $r \times c$.

$$A = \begin{vmatrix} a_{11} & . & . & . & a_{1c} \\ . & . & . & . & . \\ . & . & a_{ij} & . & . \\ . & . & . & . & . \\ a_{r1} & . & . & . & a_{rc} \end{vmatrix} \quad B = \begin{vmatrix} b_{11} & . & . & . & b_{1c} \\ . & . & . & . & . \\ . & . & b_{ij} & . & . \\ . & . & . & . & . \\ b_{r1} & . & . & . & b_{rc} \end{vmatrix}$$

La transposée matricielle

Le transposé d'une matrice A noté A^T est obtenu simplement par la permutation des positions des composants a_{ij} et a_{ji} de cette même matrice A .

La somme matricielle

La somme de deux ou plusieurs matrices de même dimension donne une nouvelle matrice (de même dimension que les matrices précédentes) dont la $(i,j)^{\text{ème}}$ composante est égale à la somme des $(i,j)^{\text{ème}}$ de chacune des matrices précédentes.

Ainsi $A+B$ donne une nouvelle matrice $E = \begin{vmatrix} e_{11} & . & . & . & e_{1c} \\ . & . & . & . & . \\ . & . & e_{ij} & . & . \\ . & . & . & . & . \\ e_{r1} & . & . & . & e_{rc} \end{vmatrix}$

dont la $(i,j)^{\text{ème}}$ composante c'est-à-dire e_{ij} donne: $e_{ij} = a_{ij} + b_{ij}$.

La différence matricielle

La différence de deux ou plusieurs matrices de même dimension donne une nouvelle matrice (de même dimension que les matrices précédentes) dont la $(i,j)^{\text{ème}}$ composante est égale à la différence $(i,j)^{\text{ème}}$ de chacune des matrices précédentes.

Ainsi $A - B$ donne une nouvelle matrice $E = \begin{vmatrix} e_{11} & . & . & . & e_{1c} \\ . & . & . & . & . \\ . & . & e_{ij} & . & . \\ . & . & . & . & . \\ e_{r1} & . & . & . & e_{rc} \end{vmatrix}$

dont la $(i,j)^{\text{ème}}$ composante c'est-à-dire e_{ij} donne: $e_{ij} = a_{ij} - b_{ij}$.

Le produit matriciel

Le produit d'une matrice par un scalaire γ donne une nouvelle matrice (de même dimension que la matrice précédente) dont la $(i,j)^{\text{ème}}$ composante est égale au produit la $(i,j)^{\text{ème}}$ composante de la matrice précédente.

Ainsi $\gamma.A$ donne une nouvelle matrice $E = \begin{vmatrix} e_{11} & . & . & . & e_{1c} \\ . & . & . & . & . \\ . & . & e_{ij} & . & . \\ . & . & . & . & . \\ e_{r1} & . & . & . & e_{rc} \end{vmatrix}$

dont la $(i,j)^{\text{ème}}$ composante c'est-à-dire e_{ij} donne: $e_{ij} = \gamma.a_{ij}$.

Le produit de deux matrices n'est possible que si le nombre de colonnes de l'une des matrices est égale au nombre de lignes de l'autre matrice.

Soient deux matrices $C = \begin{vmatrix} c_{11} & . & . & . & c_{1k} \\ . & . & . & . & . \\ . & . & c_{ij} & . & . \\ . & . & . & . & . \\ c_{m1} & . & . & . & c_{mk} \end{vmatrix}$ et $D = \begin{vmatrix} d_{11} & . & . & . & d_{1n} \\ . & . & . & . & . \\ . & . & d_{ij} & . & . \\ . & . & . & . & . \\ d_{k1} & . & . & . & d_{kn} \end{vmatrix}$

de dimensions respectives $r \times k$ et $k \times c$. Le produit de la matrice C par matrice D n'est possible que si le nombre de colonnes de la matrice C (qui est ici k) est égale au nombre de lignes de la matrice D (qui est ici k).

Ce produit $C \times D$ donne une nouvelle matrice $E = \begin{vmatrix} e_{11} & . & . & . & e_{1n} \\ . & . & . & . & . \\ . & . & e_{ij} & . & . \\ . & . & . & . & . \\ e_{m1} & . & . & . & e_{mn} \end{vmatrix}$ de

dimension $r \times c$ dont la $(i,j)^{\text{ème}}$ composante est égale à la somme des produits de la $k^{\text{ème}}$ composante de la $i^{\text{ème}}$ ligne de la matrice C et de la $k^{\text{ème}}$ composante de la $j^{\text{ème}}$ colonne de la matrice D c'est-à-dire e_{ij} donne: $e_{ij} = \sum^k a_{ik} b_{kj}$.

Cependant si $r = c$ alors le produit C.D comme le produit D.C est possible. Même si le produit C.D comme le produit D.C reste possible, force est de constater qu'en générale $C.D \neq D.C$.

La symétrie matricielle

Une matrice carrée S est qualifiée de matrice symétrique si cette matrice est égale à sa transposée c'est-à-dire $S = S^T$. Cela signifie que tous les éléments au dessus de la diagonale principale sont égaux aux éléments correspondants en dessous de la diagonale c'est-à-dire que le $(i,j)^{\text{ème}}$ élément de la matrice S est égal au $(i,j)^{\text{ème}}$ élément de sa matrice transposée S^T .

Exemple

$$A = \begin{vmatrix} 1 & 0 & 2 \\ 3 & 5 & 1 \\ 2 & 1 & 0 \end{vmatrix}$$

$$B = \begin{vmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 5 & 2 & 1 \end{vmatrix}$$

$$A + B = \begin{vmatrix} 2 & 1 & 4 \\ 3 & 5 & 2 \\ 7 & 3 & 1 \end{vmatrix}$$

$$A - B = \begin{vmatrix} 0 & -1 & 0 \\ 3 & 5 & 0 \\ -3 & -1 & -1 \end{vmatrix}$$

$$5.A = \begin{vmatrix} 5 & 0 & 10 \\ 15 & 25 & 5 \\ 10 & 5 & 0 \end{vmatrix}$$

$$A.B = \begin{vmatrix} 11 & 5 & 4 \\ 8 & 5 & 12 \\ 2 & 2 & 5 \end{vmatrix}$$

$$S = \begin{vmatrix} 1 & 5 & 0 \\ 5 & 2 & 3 \\ 0 & 3 & 1 \end{vmatrix}$$

$$S^T = \begin{vmatrix} 1 & 5 & 0 \\ 5 & 2 & 3 \\ 0 & 3 & 1 \end{vmatrix}$$

Déterminant et inverse d'une matrice carrée

Certaines opérations matricielles en analyse multivariée ne peuvent s'effectuer seulement que sur des matrices carrées.

L'une des ces opérations est l'inversion matricielle.

L'inverse d'une matrice carrée A donne une autre matrice dénotée A^{-1} dont le produit $A^{-1}.A = A.A^{-1} = I$ où I est la matrice identité.

La matrice identité est une matrice dont les composants de la diagonale principale ne sont que des 1 ($a_{ii} = 1$) et où les autres composants $a_{ij} = 0$ avec $i \neq j$.

L'inverse d'une matrice d'une matrice symétrique est aussi une matrice symétrique.

La détermination de l'inverse d'une matrice A ne peut se faire que par la détermination d'une autre quantité appelée déterminant de la matrice carrée A .

Cette quantité qualifiée de déterminant est notée $\det(A)$ ou $|A|$.

Ainsi lorsque le déterminant de la matrice carrée pour laquelle on veut déterminer l'inverse est égal à zéro alors cette matrice ne sera pas inversible.

Pour la détermination de l'inverse d'une matrice nous allons le faire que pour les matrices de dimension 2×2 ;

Soit une matrice carrée de dimension 2×2

$$A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

$$\det(A) = |A| = a_{11} \cdot a_{22} - a_{12} \cdot a_{21}$$

Si $\det(A) = |A| \neq 0$ alors A est une matrice inversible, d'inverse A^{-1} .

$$A^{-1} = \begin{vmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{vmatrix}$$

Ainsi

$$AA^{-1} = A^{-1}A = I = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

Chapitre III

THEORIE DE L'ANALYSE DISCRIMINANTE

Introduction

On commencera par la présentation de quelques exemples qui nous permettrons de dégager les caractéristiques d'un problème de discrimination.

Dans une banque on peut s'interroger sur le fait de savoir si un emprunteur potentiel représente un bon ou un mauvais risque. On peut, pour se faire, utiliser des données individuelles telles que le revenu, l'âge, la catégorie socio-professionnelle. Ce qui suppose obligatoirement que la banque possède des informations sur sa clientèle précédent.

En anthropologie, on peut s'interroger sur le sexe d'un individu dont on a découvert le crâne lors des fouilles archéologiques. Là aussi, les anthropologues pour déterminer ce sexe, pourront utiliser des données anthropométriques d'échantillons de crânes que l'on sait provenir d'hommes ou de femmes.

Le point commun à tous ces exemples peut se schématiser comme suit: on dispose de catégories ou de groupes dans lesquels il faut classer un nouvel élément soumis à l'analyse.

Il apparaît clairement dans les exemples que le but de l'analyse discriminante est de classer un individu dans un groupe sur base du profil observé de ses caractéristiques.

Cette analyse présuppose l'existence de groupes d'individus bien définis. Ces groupes peuvent être formés dès le départ sur bases de considérations naturelles mais peuvent parfois résulter d'une analyse préliminaire.

L'analyse discriminante proprement dite ne s'intéresse pas au problème de détermination des groupes. Elle est donc à distinguer d'autres méthodes d'analyse multivariée comme les méthodes de classification qui cherchent à mettre en évidence certains groupements qui se dégagent des observations. En analyse discriminante, on admet l'existence d'un nombre fixé g de groupes. Dans le cas où les caractéristiques des divers groupes ne sont pas parfaitement connues, on supposera généralement l'existence d'échantillons représentatifs de ceux-ci.

De façon formelle, un individu sera caractérisé par un vecteur x de R^p , dont les composantes peuvent être continues ou discrètes.

Les groupes seront généralement caractérisés par une distribution de probabilité. On notera $f_k(x)$ la densité relative à la distribution des p variables au sein du groupe k .

Nous utiliserons la notation X_k pour représenter la matrice ($n_k \times p$) des n_k observations de l'échantillon du groupe k .

Une règle de discrimination revient à diviser l'espace R_p en g régions notées R_1, R_2, \dots, R_g telles que si le vecteur associé à un nouvel individu appartient à R_k , cet individu est classé dans le groupe k .

I - Tests de laboratoire clinique

Les tests de laboratoire clinique sont des examens chimiques sur lesquels le clinicien se base pour prédire si un sujet donné est malade ou pas .

On distingue deux grands types de tests:

- les tests binaire ou qualitatif qui permet de spécifier la présence ou l'absence d'une pathologie(maladie)
- les tests quantitatifs permettent de spécifier en valeur réelle des mesures effectuées sur un patient.

On distingue dans cette catégorie les tests continus et les tests discrets.

Les tests continus sont des tests qui peuvent prendre des valeurs réelles alors que les tests discrets sont qui ne prennent que des valeurs entières

Un test laboratoire quelque soit son type doit satisfaire deux conditions, à savoir:

- validité: on lui reconnaît des qualités suffisantes pour être interprété en vue d'une décision médicale. Cette validité est appréciée sur sa capacité à isoler les sujets malades et les sujets sains.
- Reproductibilité: la répétition du même test chez un même sujet, ayant le même état doit donner le même résultat

II - Formulation du problème

Supposons que la population étudiée soit subdivisée en g sous-populations distinctes et bien définies.

Un individu extrait de cette population lui fournit une observation $x^t = x' = (x_1, \dots, x_p)$ de p variables le caractérisant.

Le problème qui se pose à nous est alors de classer cet individu sur base de l'observation x , dans une et une seule des g sous populations.

Ce qui implique que l'analyse discriminante se présente sous forme d'un problème de décision.

L'ensemble A des actions possibles est donc $A = \{a_j: 1 \leq j \leq g\}$ où l'action a_j consiste à classer l'individu dans la sous-population j . Remarquons qu'il existe une bijection entre cet ensemble et l'ensemble des états possibles de l'individu $\Pi = \{\Pi_k: 1 \leq k \leq g\}$ où l'état Π_k signifie que l'individu appartient à la sous population k .

On supposera que chacune des g sous-population peut être spécifiée au moyen d'une probabilité P_k définie sur les événements relatifs à un vecteur aléatoire $X' = (X_1, \dots, X_p)$ dont les composantes représentent les mesures aléatoires des p différentes caractéristiques.

Deux cas peuvent être envisagés:

- le cas où X est un vecteur aléatoire continue pour lequel les densités de probabilité $f_k(x)$ sont définies par

$$f_k(x) = \frac{\partial}{\partial x_1} \dots \frac{\partial}{\partial x_p} P_k(X_1 \leq x_1, \dots, X_p \leq x_p)$$

- le cas où X est un vecteur aléatoire discret pour lequel les densités de probabilités $f_k(x)$ sont définies par

$$f_k(x) = P_k(X_1 = x_1, \dots, X_p = x_p)$$

Si S désigne l'ensemble des valeurs possibles de x et si f est une fonction réelle définie sur S , l'espérance par rapport à la probabilité P_k de f_k sera fournie dans le cas continu et dans le cas discret respectivement par

$$E_k f(X) = \int_S f(x) f_k(x) \quad (II.1).$$

et par

$$E_k f(X) = \sum_S f(x) f_k(x) \quad (II.2)$$

Pour des raisons de simplicité, nous n'envisagerons, dans la suite, que le cas continu. Tout ce qui sera dit sera également valable dans le cas discret moyennant la transformation des formules du type (II.1) en formule du type (II.2).

III - Règles de décision

Il existe deux types de règles de décision: les règles de décision pures ou non-aléatoires et les règles de décision aléatoires.

Une règle de décision pure assigne à chaque valeur de $x \in S$, une action de A . Par conséquent, si S_j désigne l'ensemble des x pour lesquels la décision est l'action a_j , une règle de décision pure est entièrement caractérisée par la partition de S en $\{S_j: 1 \leq j \leq g\}$. Ainsi un individu sera classé dans la population j si $x \in S_j$.

Une règle de décision aléatoire, d , fournit, pour chaque valeur de $x \in S$, un vecteur $d(x) = \{d_j(x): 1 \leq j \leq g\}$ tel que $d_j(x) \geq 0 \quad \forall j$ et $\sum_{1 \leq j \leq g} d_j(x) = 1$.

Signalons que les règles de décision pures ne sont, en fait, que des cas particuliers de règles de décision aléatoires. En effet, si on désigne par I_A la fonction indicatrice d'un ensemble A , et si l'on considère la règle de décision aléatoire d telle que $d_j(x) = I_{S_j(x)} \quad \forall j$, un individu fournissant l'observation $x \in S_j$ sera presque sûrement classé dans la population j .

Si pour une règle de décision aléatoire d , on désigne par $\epsilon_{jk}(d)$ la probabilité qu'un individu provenant de la population k (Π_k) soit classé par d dans la population j (a_j), c'est-à-dire

$$\epsilon_{jk}(d) = E_k[d_j(X)]$$

ou encore dans le cas continu

$$\epsilon_{jk}(d) = \int_S d_j(x) f_k(x) dx \quad (III.1)$$

Dans le cas particulier d'une règle de décision pure, on aura

$$\epsilon_{jk}(d) = P_k(X \in S_j) = \int_{S_j} f_k(x) dx$$

On se doit de rechercher une règle de décision qui nous permet de contrôler les erreurs provenant du mauvais classement dans la population j d'un individu. Pour cela nous allons faire recours dans notre analyse à une fonction de perte $L_{jk} = L[a_j, \Pi_k]$ qui représentera le coût associé au mauvais classement dans la population j d'un individu provenant de la population k . On supposera qu'il n'y a pas de récompense pour le bon classement d'un individu, c'est-à-dire $L_{jj} = 0 \forall j$. Généralement, on peut considérer que les coûts sont identiques, mais il n'est pas difficile de concevoir des situations où certaines erreurs sont bien plus graves que d'autres. La qualité d'une règle de décision d sera appréciée par l'espérance du coût de mauvaise classification d'un individu provenant de la population k pour les différents valeurs de k , c'est-à-dire le vecteur des risques inhérents à la règle de décision d , soit $R(d) = \{R_k(d) : 1 \leq j \leq g\}$.

En fait

$$R_k(d) = E_k[L(D, \Pi_k)]$$

$$R_k(d) = \sum_{1 \leq j \leq g} L_{jk} \epsilon_{jk}(d) \quad (III.2)$$

Remarquons que dans le cas particulier où $L_{jk} = L \forall j \neq k$, $R_k(d) = L[1 - \epsilon_{kk}(d)]$ où $\epsilon_{kk}(d)$ est la probabilité de classer correctement un individu provenant de la population k au moyen de la règle de décision d .

Comme il est difficile sinon presque impossible de déterminer une règle de décision optimale qui puisse minimiser à la fois tous les risques alors il devient indispensable soit de posséder plus de renseignements sur les g populations, soit de définir des critères additionnels permettant de cerner ce choix. Ainsi par exemple si l'on connaît les probabilités a priori qu'un individu extrait de la population appartienne aux différentes sous-populations, c'est-à-dire $p = \{P_k : 1 \leq k \leq g\}$, il est possible de définir l'espérance du coût total de mauvaise classification d'un individu extrait de la population, qui est appelée risque bayésien d'une règle de décision d et que l'on dénote par $R(d, p)$.

En fait,

$$R(d,p) = \sum_{1 \leq k \leq g} P_k R_k(d).$$

On peut chercher une règle de décision minimisant le risque bayésien et une telle règle de décision, d , sera appelée règle de décision bayésienne correspondant à p et sera caractérisée par le fait que pour tout autre règle de décision aléatoire d' ,

$$R(d,p) \leq R(d',p)$$

Le théorème suivant donne la forme générale des règles de décision bayésiennes.

Théorème

Toute règle de décision bayésienne (correspondant à p) est de la forme

$$\forall x \in S \quad d_j(x) = 0 \quad \text{si} \quad s_j(x) < s(x)$$

où

$$s_j(x) = - \sum_{1 \leq k \leq g} L_{jk} P_k f_k(x)$$

$$s(x) = \max_{1 \leq j \leq g} s_j(x)$$

De plus, le risque bayésien (minimal) correspondant est donnée, dans le cas continu, par

$$R(d,p) = - \int_S s(x) dx$$

Démonstration: En se référant aux formules (III.1) et (III.2), il est facile de voir que $R(d,p)$ peut s'écrire sous la forme

$$R(d,p) = - \int_S \sum_{1 \leq j \leq g} s_j(x) d_j(x) dx$$

et, par conséquent, puisque $\sum_{1 \leq j \leq g} d_j(x) = 1$ alors

$$R(d,p) + \int_S s(x) dx = \sum_{1 \leq j \leq g} \int_S \{s(x) - s_j(x)\} d_j(x) dx$$

Le second membre de cette égalité est toujours plus grand ou égal à zéro et ne sera égal à zéro pour tout $x \in S$ que si $d_j(x) = 0$ car $s(x) - s_j(x) > 0$.

On appelle $s_j(x)$ le score discriminant pour la population j d'un individu caractérisé par x .

Il résulte de ce théorème que toute règle de décision bayésienne peut être décrite de la manière suivante: si un individu extrait de la population fournit l'observation x , il sera classé au hasard dans l'une des populations j telles que $s_j(x) = \max_{1 \leq j' \leq g} s_{j'}(X)$

Dans le cas particulier où $L_{jk} = 1, \forall j \neq k$, il est facile de voir que

$$s_j(x) = L[P_j f_j(x) - f(x)]$$

$$\text{où } f(x) = \sum_{1 \leq j \leq g} P_j f_j(x)$$

De là, les règles de décision bayésiennes seront, dans ce cas, de la forme suivante:

$$\forall x \in S \quad d_j(x) = 0 \quad \text{si} \quad p_j f_j(x) < \max_{1 \leq j' \leq g} P_{j'} f_{j'}(X)$$

c'est-à-dire qu'un individu extrait de la population et fournissant l'observation x sera classé au hasard dans l'une des populations j telles que $P_j f_j = \max_{1 \leq j' \leq g} P_{j'} f_{j'}(X)$

Une telle règle de décision est appelée règle de classification du rapport de vraisemblance et, lorsque $P_j = \frac{1}{g} \quad \forall j$, elle est appelée règle du maximum de vraisemblance.

Si l'on envisage, à présent, le cas particulier où l'on ne considère que deux sous-populations ($g = 2$), on voit aisément que:

$$s_1(x) = -L_{12} P_2 f_2(x)$$

$$s_2(x) = -L_{21} P_1 f_1(x)$$

et, par conséquent, la règle de décision bayésienne sera fournie par:

$$d(x) = \begin{vmatrix} 1 \\ 0 \end{vmatrix} \quad \text{si} \quad \frac{f_1(x)}{f_2(x)} > c$$

$$d(x) = \begin{vmatrix} 0 \\ 1 \end{vmatrix} \quad \text{si} \quad \frac{f_1(x)}{f_2(x)} < c$$

$$\text{où } c = \frac{L_{12} P_2}{L_{21} P_1}$$

c'est-à-dire le rapport des risques bayésiens correspondant aux actions a_1 et a_2 .

IV - Méthodes de décision bayésiennes en analyse discriminante

Il est utile de distinguer en statistique d'une part les procédures "descriptives" et d'autre part les procédures inférentielles.

Les procédures descriptives se donnent pour tâche de résumer des distributions (de probabilité ou de fréquence) au moyen de caractéristiques diverses. Souvent ces caractéristiques sont évaluées pour leur seul intérêt descriptif, c'est-à-dire sans référence explicite à un modèle d'échantillonnage.

Les procédures inférentielles se basent sur un "modèle statistique", c'est-à-dire sur un ensemble de distribution de probabilité.

On distingue deux courants dans cette procédure qui sont les procédures "classiques" ou "d'échantillonnage" et les procédures dites "bayésiennes".

L'observation analysée x est alors considérée comme la réalisation d'une distribution. Il sera commode d'indexer l'ensemble de ces distributions par un paramètre θ . On écrira alors le modèle M de la façon suivante:

$$M = \{p(x | \theta) = f(x | \theta) : \theta \in \Theta\}$$

où Θ est l'espace paramétrique. L'idée importante est que à tout θ correspond une et une seule distribution de probabilité

Exemple: On suppose que des observation x_1, x_2, \dots, x_n ont été engendrées de façon indépendante et que chaque observation x_i a été engendrée par une distribution normale de moyenne μ et de variance σ^2 . Dans l'écriture du modèle M ci-dessus, $p(x | \theta)$ aura la forme:

$$P(x|\theta) = \prod_{i=1}^n P(x_i | \theta)$$

$$P(x_i|\theta) = (2\pi)^{-\frac{1}{2}} s^{-1} \text{Exp} \frac{-1}{2} \left(\frac{x_i - m}{s} \right)^2$$

$$\theta = (\mu, \sigma^2) \quad \Theta = \mathbb{R} \times \mathbb{R}_+$$

Dans cet exemple, la connaissance de θ est suffisante pour caractériser la distribution $p(x | \theta)$.

Dans une optique "classique" les procédures statistiques sont analysées en tant que fonction de la variable aléatoire x sujette à une des distributions $p(x | \theta)$. Les propriétés retenues pour la comparaison et le choix de ces procédures sont des propriétés "d'échantillonnage". Typiquement pour une procédure donnée (d'estimation ponctuelle, d'intervalle de confiance ou de test d'hypothèse), ces propriétés pourront être "bonnes" pour certaines valeurs de θ et "mauvaises" pour d'autres valeurs de θ .

Dans une optique "bayésienne", on considère que l'essence de l'inférence statistique consiste à "apprendre en observant", et on considère que cet apprentissage est caractérisé par la révision d'une distribution de probabilité dite "à priori" $p(\theta)$ en une distribution $p(x | \theta)$ dite distribution "à postérieure". Le modèle statistique bayésien consiste en une et une seule distribution de probabilité $p(x, \theta)$ portant sur le paramètre et l'observation.

Cette distribution est formée par une distribution à priori $p(\theta)$ et les distributions conditionnelles $p(x | \theta)$ représentant le processus d'échantillonnage:

$$p(x, \theta) = p(\theta).p(x | \theta)$$

L'objet des calculs statistiques est d'opérer la décomposition duale de cette distribution conjointe:

$$p(x, \theta) = p(\theta).p(\theta | x)$$

La distribution marginale sur les observation, $p(x)$ est appelée distribution "prédictive" et la distribution conditionnelle sur les paramètres, $p(\theta | x)$ est appelée distribution à posteriori. En résumé, un modèle statistique bayésien s'écrit donc sous la forme d'une seule distribution de probabilité, $p(x, \theta)$, décomposable de deux façons différentes:

$$M = \{p(x, \theta) = p(\theta).p(x | \theta) = p(x). p(\theta | x)\}.$$

Le calcul des distributions prédictives $p(x)$, de la probabilité à posteriori $p(\theta | x)$, et de leurs caractéristiques principales à partir des distributions à priori $p(\theta)$ et d'échantillonnage $p(x | \theta)$, forme l'essentiel des calculs propres aux méthodes bayésiennes. Ces calculs se baseront dans le cas continu, sur les formules suivantes:

$$p(x) = \int p(\theta).p(x | \theta)d\theta$$

$$p(\theta | x) = \frac{p(\theta).p(x|\theta)}{p(x)}$$

La distribution prédictive peut être vue comme une "moyenne" des distributions d'échantillonnage $p(x | \theta)$, moyenne pondérée par la distribution a priori $p(\theta)$.

Nous savons que l'analyse discriminante se présente naturellement sous la forme d'un problème de décision. Ici on envisagera seulement que les règles de décision non aléatoires. Une règle de décision sera écrite " a_j " ($1 \leq j \leq g$) au lieu de d ; la distribution à priori sur les g populations sera toujours considérée comme fixée une fois pour toutes mais sera sujette à révision après avoir observé une variable x . Dès lors, on utilisera la fonction de perte $L_{jk} = L(a_j, \Pi_k)$ qui représente le coût associé à la décision a_j de classer dans la population Π_j alors que la population réelle était Π_k . Le risque bayésien devient la perte attendue, c'est-à-dire l'espérance mathématique de la perte L_{jk} par rapport à une distribution sur les g population Π_j ; si on utilise la distribution à priori, on n'écrit simplement $R(a_j)$ (au lieu de $R(d, p)$) et, s'il s'agit de la distribution à posteriori, après avoir observé x , on écrira $R(a_j | x)$ (au lieu de $R(d, p)$ où d sera la règle non aléatoire a_j et p la distribution à posteriori).

Dans une optique de comportement rationnel face à l'incertitude, on représentera les opinions au sujet de la vraisemblance des différents états de la nature au moyen d'une distribution de probabilités "à priori" $p = (p_1, p_2, \dots, p_g)$ où $p_k = p(\Pi_k)$ représente la probabilité que nous assignons à priori à la proposition " x provient de la population Π_k ".

L'objectif est de minimiser le risque à priori.

$$R(a_j) = EL(a_j, \Pi) = \sum_k p(\Pi_k) L_{jk}$$

Avant donc d'observer la valeur de x , la décision optimale est donnée par:

$$\min_j R(a_j) = \min_j \sum_k p(\Pi_k) L_{jk}$$

Lorsque la valeur de x est observée, on peut alors réviser son opinion à priori et calculer la probabilité à posteriori $p(\Pi_k | x)$ par la formule usuelle:

$$p(\Pi_k | x) = \frac{p(\Pi_k) p(x | \theta_k)}{p(x)}$$

où

$$p(x) = \sum_k p(\Pi_k) p_k(x | \theta_k)$$

c'est la densité prédictive qui permet de prédire la valeur de x si on ignore de quelle population elle est tiré. On calcule alors le risque à posteriori de chaque décision a_j :

$$R(a_j | x) = \sum_k p(\Pi_k | x) L_{jk}$$

la procédure optimale devient la règle de décision du type:

“Assigner l'observation l'observation x à la population Π_j si $R(a_j | x) \leq R(a_i | x)$ ($i = 1, 2, \dots, g$)”

On définit, dès lors, une fonction $x \rightarrow d(x) = a_j$ dont on peut calculer le risque bayésien, c'est-à-dire la perte attendue avant d'avoir effectivement observé x :

$$\begin{aligned} R(d) &= E R(d(x) | x) = \int R(d(x) | x) p(x) dx \\ &= \sum_k p(\Pi_k) \int L(d(x), \Pi_k) p_k(x | \theta_k) dx \end{aligned}$$

Remarquons que toute règle de décision définit une partition de l'espace des x en g régions R_1, R_2, \dots, R_g par la règle “ $d(x) = a_j$ ssi $x \in R_j$ ”.

En examinant de près la situation deux populations ($g = 2$), le problème revient à déterminer deux régions R_1 et R_2 qui minimisent la perte attendue. Une règle de décision d sera représentée par (R_1, R_2) et son risque bayésien sera:

$$\begin{aligned} R(d) &= R(R_1, R_2) = E[L(a, \Pi)] \\ &= L_{21} p_1 p(x \in R_2 | \theta_1) + L_{12} p_2 p(x \in R_1 | \theta_2) \\ &= L_{21} p_1 \int_{R_2} f(x | \theta_1) dx + L_{12} p_2 \int_{R_1} f(x | \theta_2) dx \end{aligned}$$

Un calcul simple montre que la partition (R_1, R_2) qui minimise cette expression est obtenue de la façon suivante:

$$R_1 = \{x | L_{12} p_2 f(x | \theta_2) < L_{21} p_1 f(x | \theta_1)\}$$

$$R_2 = \{x | L_{12} p_2 f(x | \theta_2) \geq L_{21} p_1 f(x | \theta_1)\}$$

Plus généralement, lorsque le nombre de groupe g est plus grand que 2 ($g > 2$), la partition optimale sera obtenue de la façon suivante:

$$R_j = \left\{ x \mid \sum_{\substack{k=1 \\ k \neq j}}^g p_k L_{jk} f_k(x | \theta_k) < \min_L \sum_{\substack{k=1 \\ k \neq L}}^g p_k L_{Lk} f_k(x | \theta_k) \right\}$$

A titre d'exemple, considérons le cas de deux populations normales:

$$f_k(x | \theta_k) = (2\pi)^{-\frac{1}{2}} \left| \sum_k \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_k)' \sum_k^{-1} (x - \mu_k) \right\} \quad k = 1, 2$$

Le paramètre θ_k est donc $\theta_k = (\mu_k, \sum_k)$. on obtient alors la règle de classification suivante:

$$R_1 = \left\{ x \mid L_{12} p_2 \left| \sum_2 \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_2)' \sum_2^{-1} (x - \mu_2) \right\} < L_{21} p_1 \left| \sum_1 \right|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_1)' \sum_1^{-1} (x - \mu_1) \right\} \right\}$$

ou encore

$$R_1 = \left\{ x \mid (x - \mu_2)' \sum_2^{-1} (x - \mu_2) - (x - \mu_1)' \sum_1^{-1} (x - \mu_1) > 2 \log \left(\frac{\left| \sum_1 \right|^{-\frac{1}{2}} p_2 L_{12}}{\left| \sum_2 \right|^{-\frac{1}{2}} p_1 L_{21}} \right) \right\}$$

On obtient donc une région R_1 définie par:

$$R_1 = \{x | x'Ax + b'x + c > d\}$$

où

$$A = \Sigma_2^{-1} - \Sigma_1^{-1}$$

$$b = -2 \left(\Sigma_2^{-1} \mu_2 - \Sigma_1^{-1} \mu_1 \right)$$

$$c = \mu_2' \Sigma_2^{-1} \mu_2 - \mu_1' \Sigma_1^{-1} \mu_1$$

$$d = 2 \text{Log} \left(\frac{|\Sigma_1|^{-\frac{1}{2}} p_2 L_{12}}{|\Sigma_2|^{-\frac{1}{2}} p_1 L_{21}} \right)$$

La fonction quadratique en x " $x'Ax + b'x + c$ " est qualifiée de fonction discriminante.

Lorsque les matrices de variance-covariance et les coûts des erreurs de classification sont les mêmes (c'est-à-dire $\Sigma_1 = \Sigma_2 = \Sigma$ et $L_{21} = L_{12} = L$), on obtient une région R_1 définie par:

$$R_1 = \{x | a_1'x > a_2\}$$

où

$$a_1 = \Sigma^{-1} (\mu_1 - \mu_2)$$

$$a_2 = \frac{1}{2} (\mu_1' \Sigma^{-1} \mu_1 - \mu_2' \Sigma^{-1} \mu_2)$$

La fonction linéaire $a_1'x$ est la fonction de discrimination.

Le problème de l'analyse discriminante a été abordé sous plusieurs approches par les spécialistes dont la plus connue est la discrimination classique ou linéaire. La discrimination linéaire est une approche est plutôt descriptive car elle ne repose sur aucune hypothèse spécifique de distribution. Les concepts utilisés sont de nature géométrique (distance, métriques de discrimination, axes factoriels discriminants).

A côté de cette approche classique d'autres approches ont été développées comme l'approche logistique qui ne tient pas compte des hypothèses spécifiques des distributions mais des propriétés bayésiennes.

La discrimination logistique présente l'avantage d'être applicable dans un cadre relativement général et permet une estimation directe de la fonction discriminante caractérisant la règle de discrimination.

V - Discrimination linéaire

Nous allons présenter l'analyse discriminante linéaire en ne considérant que deux groupes. Chaque groupe k ($k=1,2$) étant caractérisé par un ensemble de n_k observations de p variables. On disposera pour chaque groupe k , d'une matrice X_k de format $n_k \times p$. on désigne par X la matrice de format $n \times p$ formée par les 2 matrices X_k

Pour un individu dont on ignore le groupe et qui est caractérisé par un vecteur $x \in \mathbb{R}^p$, il s'agit d'affecter cet individu à un des 2 groupes.

A) Décomposition de la matrice de dispersions

Si on note \bar{x} le vecteur des moyennes des p variables pour l'ensemble des n individus, et \bar{x}_j le vecteur des moyennes de ces mêmes variables pour les n_j individus du groupe j , la matrice A des dispersions totales s'écrit:

$$A = X'X - n \bar{x} \bar{x}'$$

et les matrices de dispersion dans les groupes s'écrivent

$$A_k = X_k' X_k - n_k \bar{x}_k \bar{x}_k' \quad k=1, 2$$

Dès lors,

$$A = \sum_{1 \leq k \leq 2} X_k' X_k - n \bar{x} \bar{x}'$$

$$A = \sum_{1 \leq k \leq 2} (A_k + n_k \bar{x}_k \bar{x}_k') - n \bar{x} \bar{x}'$$

$$A = \sum_{1 \leq k \leq 2} A_k + \sum_{1 \leq k \leq 2} n_k (\bar{x}_k \bar{x}_k' - \bar{x} \bar{x}')$$

de sorte que la matrice A peut s'écrire

$$A = \sum_{1 \leq k \leq 2} A_k + \sum_{1 \leq k \leq 2} n_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})'$$

ou encore

$$A = W + B \quad (V.1)$$

et se trouve décomposée en une somme de deux matrices W et B ; la matrice W , qui est la somme des matrices de dispersion des p variables au sein des groupes, sera appelée matrice de dispersion intra-groupes, et la matrice B , qui est la matrice de dispersion des p variables pour les 2 centres de groupes, chaque centre étant affecté d'un poids égal à l'effectif du groupe, sera appelée matrice de dispersion inter-groupes.

Cette matrice de dispersion inter-groupe s'écrivant de la façon suivante

$$B = \sum_{k=1}^2 n_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})'$$

$$B = n_1 (\bar{x}_1 - \bar{x})(\bar{x}_1 - \bar{x})' + n_2 (\bar{x}_2 - \bar{x})(\bar{x}_2 - \bar{x})'$$

en utilisant la relation

$$\bar{x} = \frac{1}{n} (n_1 \bar{x}_1 + n_2 \bar{x}_2)$$

on obtient l'expression

$$B = \frac{n_1 n_2}{n} (\bar{x}_1 - \bar{x}_2)(\bar{x}_1 - \bar{x}_2)'$$

Cette expression montre que la matrice B est le produit d'une matrice colonne M par sa transposée M'. La matrice M étant définie par l'expression

$$M = \sqrt{\frac{n_1 n_2}{n}} (\bar{x}_1 - \bar{x}_2)$$

d'où on peut écrire l'expression de B de la façon suivante

$$B = M M'$$

B) Détermination des axes factoriels discriminants

A chaque individu i est associé un vecteur $x_i \in R^p$, de sorte que les n individus peuvent être considérés comme n points de l'espace $\in R^p$.

On recherche une direction $c \in R^p$ telle que, d'après la projection sur c, les 2 groupes soient "discriminés" le mieux possible. Dans ce but, on considère le vecteur $u = X c \in R^n$ formé des projections $x_i' c$ des n points x_i sur c: c'est le vecteur des valeurs prises par la variable c'x chez chacun des n individus.

La dispersion d du vecteur u s'écrit, si \bar{u} désigne la moyenne des composantes de u et i le vecteur de R^n dont toutes les composantes sont 1,

$$d = u' u - n \bar{u}^2$$

$$d = u' u + \frac{-1}{n} u' i i' u$$

ou encore, en remplaçant u par Xc ,

$$d = c' A c$$

En utilisant (V.1), on obtient la décomposition

$$d = c' W c + c' B c$$

de la dispersion de u en une dispersion intra-groupes $c' W c$ et une dispersion inter-groupes $c' B c$.

Pour discriminer au mieux les projections des groupes sur c , il est naturel de choisir le vecteur c qui maximise la dispersion inter-groupes et qui minimise la dispersion intra-groupes. On choisira donc un vecteur c qui maximise le quotient $\frac{c' B c}{c' W c}$ ou encore, ce qui est équivalent, qui à minimiser $1 + \frac{c' W c}{c' B c}$. Cette dernière expression étant égale à $\frac{c' A c}{c' B c}$, on est conduit à maximiser la fonction homogène

$$f(c) = \frac{c' B c}{c' A c}$$

Si on considère une droite de R^p issue de l'origine, la valeur de f est la même en tout point distinct de l'origine de cette droite. En conséquence, on peut se borner à rechercher le maximum de f sur l'ellipsoïde d'équation $c' A c = 1$, c'est-à-dire à rechercher le maximum de

$$h(c) = c' B c$$

sous la contrainte

$$c' A c = 1.$$

La dérivation du lagrangien

$$c' B c - \lambda (c' A c - 1)$$

livre la relation

$$B c = \lambda A c$$

qui devient, si A est inversible,

$$A^{-1} B c = \lambda c \quad (V.2)$$

Le vecteur c recherché est donc le vecteur propre de $A^{-1}B$ correspondant à la plus grande valeur propre λ de $A^{-1}B$.

En remplaçant B par sa valeur, soit $B = M M'$, on a la relation (V.2) qui devient

$$A^{-1} M M' c = \lambda c$$

où encore, après multiplication à gauche par M' , cela donne

$$(M' A^{-1} M) M' c = \lambda M' c.$$

De la relation ci-dessus on déduit

$$\lambda = M' A^{-1} M$$

Le vecteur propre c est appelé l'axe canonique ou l'axe factoriel et la valeur propre λ qui lui correspond est appelée pouvoir discriminant de la variable $c'x$ associée. Cette variable $c'x$ est appelée la variable canonique ou la variable discriminante.

Le pouvoir discriminant λ est égal, à une constante près, à la distance entre les centres de gravités des deux groupes pour la métrique définie par S^{-1} . Au coefficient $\frac{n_1 n_2}{n^2}$ près, ce pouvoir discriminant n'est rien d'autre que la distance de Mahalanobis

$$D^2 = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2)$$

c'est la distance entre les centres de classes pour la métrique définie par S^{-1} . Le vecteur propre c correspondant est la solution de

$$M M' c = \lambda A c$$

soit le vecteur

$$c = S^{-1} M$$

La valeur λ est donc toujours comprise entre 0 et 1. Elle est égale à 1 si la dispersion intra-groupes est nulle, c'est-à-dire si les points d'un même groupe se projettent en un même point de c ; elle est égale à 0 si la dispersion inter-groupes est nulle, c'est-à-dire si les centres de chacun des groupes se projettent tous en un même point de c .

C) Affectation d'un individu à un groupe

Supposons qu'on dispose des vecteurs x_i relatifs à n individus parmi lesquels n_1 appartiennent au groupe 1 et n_2 au groupe 2.

A partir de ces vecteurs x_i , on peut construire la matrice A et les matrices A_k et en déduire une procédure d'affectation à un des 2 groupes pour un individu caractérisé par un vecteur x et dont on ignore le groupe auquel il appartient; on calcule pour cela, la distance de x à chacun des 2 groupes à l'aide de la formule:

$$Dis(x, x_k) = |S_k|^{-\frac{1}{p}} (p + (x - \bar{x}_k)' S_k^{-1} (x - \bar{x}_k))$$

où \bar{x}_k est le centre de gravité du groupe k et S_k la matrice $\frac{1}{n_k}A_k$ des variances-covariances de ce groupe; on détermine la plus petite distance de ces distances et on affecte à l'individu x au groupe correspondant.

La méthode d'affectation qui vient d'être décrite nécessite le calcul des déterminants et des inverses de chacune des matrices S_k . Ceci explique qu'on se contente souvent d'adopter pour distance entre x et X_k l'expression

$$|S|^{-1/2} (p + (x - \bar{x}_k)' S^{-1} (x - \bar{x}_k))$$

où $S = \frac{1}{n}A$, ce qui revient à considérer que les dispersions sont homogènes dans les groupes.

Il suffit alors de comparer les nombres

$$(x - \bar{x}_k)' S^{-1} (x - \bar{x}_k) \quad k = 1, 2$$

le terme $x' S^{-1} x$ étant commun, il reste à comparer les nombres

$$(\bar{x}_k - 2x)' S^{-1} \bar{x}_k \quad k = 1, 2$$

On décidera d'affecter un individu caractérisé par un vecteur x au groupe 1 si

$$(\bar{x}_1 - 2x)' S^{-1} \bar{x}_1 < (\bar{x}_2 - 2x)' S^{-1} \bar{x}_2$$

ou si

$$\bar{x}_2' S^{-1} \bar{x}_2 - \bar{x}_1' S^{-1} \bar{x}_1 + 2x' S^{-1} \bar{x}_1 - 2x' S^{-1} \bar{x}_2 > 0$$

ou encore si

$$\bar{x}_2' S^{-1} \bar{x}_2 - \bar{x}_1' S^{-1} \bar{x}_1 + 2x' S^{-1} (\bar{x}_1 - \bar{x}_2) > 0$$

On remarque que la somme des deux premiers termes de cette expression peut s'écrire

$$(\bar{x}_1 + \bar{x}_2)' S^{-1} (\bar{x}_2 - \bar{x}_1)$$

on obtient, après division par 2, la relation

$$x' S^{-1} (\bar{x}_1 - \bar{x}_2) > \frac{1}{2} (\bar{x}_1 + \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2)$$

qui s'écrit encore

$$(x - \frac{1}{2}(\bar{x}_1 + \bar{x}_2))' S^{-1} (\bar{x}_1 - \bar{x}_2) > 0 \quad (V.3)$$

$S^{-1}(\bar{x}_1 - \bar{x}_2)$ est, à une constante positive près, l'axe discriminant c.

La relation (V.3) implique qu'on affecte l'individu x à la classe 1 si la projection du vecteur joignant le point milieu des centres \bar{x}_1 et \bar{x}_2 au point x sur l'axe discriminant est positive

D) Exemple

Ce tableau ci-dessous donne les valeurs centrées des variables x_1 et x_2 chez 24 individus répartis en deux classes.

CLASSE 1		CLASSE 2	
-5	1	-3	0
-4	-1	-2	-2
-4	3	-1	2
-3	-2	-1	4
-3	2	1	-3
-1	-4	1	-1
-1	-2	1	3
-1	0	2	1
1	-3	3	4
2	-4	4	-1
2	-1	4	2
5	-1	3	3
-12	-12	12	12

On recherche d'abord la matrice des covariances

$$S = \frac{1}{24} \begin{bmatrix} -5 & -4 & \dots & 4 & 3 \\ 1 & -1 & \dots & 2 & 3 \end{bmatrix} \begin{bmatrix} -5 & -1 \\ -4 & -1 \\ \cdot & \cdot \\ \cdot & \cdot \\ 4 & 2 \\ 3 & 3 \end{bmatrix} = \frac{1}{24} \begin{bmatrix} 184 & -1 \\ -1 & 140 \end{bmatrix}$$

puis la matrice S^{-1}

$$S^{-1} = \frac{24}{25759} \begin{bmatrix} 140 & 1 \\ 1 & 184 \end{bmatrix} = \begin{bmatrix} 0.13044 & 0.00093 \\ 0.00093 & 0.17112 \end{bmatrix}$$

et le vecteur M est donnée par

$$S^{-1} = \sqrt{6} (\bar{X}_1 - \bar{X}_2) = \sqrt{6} \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} = \sqrt{6} \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

on en déduit le vecteur $C = S^{-1}M$ qui le seul axe factoriel discriminant

$$C = \sqrt{6} \begin{bmatrix} 0.13044 & 0.00093 \\ 0.00093 & 0.17112 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \sqrt{6} \begin{bmatrix} -0.26272 \\ -0.34410 \end{bmatrix}$$

ou encore, en normant ce vecteur C, $C = (-0.6069, -0.7948)$.

La variable discriminante est donc la variable

$$u = [-0.6069 \quad -0.7948]x$$

La figure 1 ci-dessous nous donne la position de l'axe factoriel C dans le nuage des points centrés et la figure 2 donne les valeurs u_i de la variable u chez chacun des 24 individus.

Un individu quelconque sera affecté au groupe 1 si $[-0.6069 \quad -0.7948]x \geq 0$ ou au groupe 2 sinon.

On observe sur cet exemple que sur les 24 individus seulement 6 individus ont connu une mauvaise affectation dont 4 en provenance de la classe 2 et 2 en provenance de la classe 1.

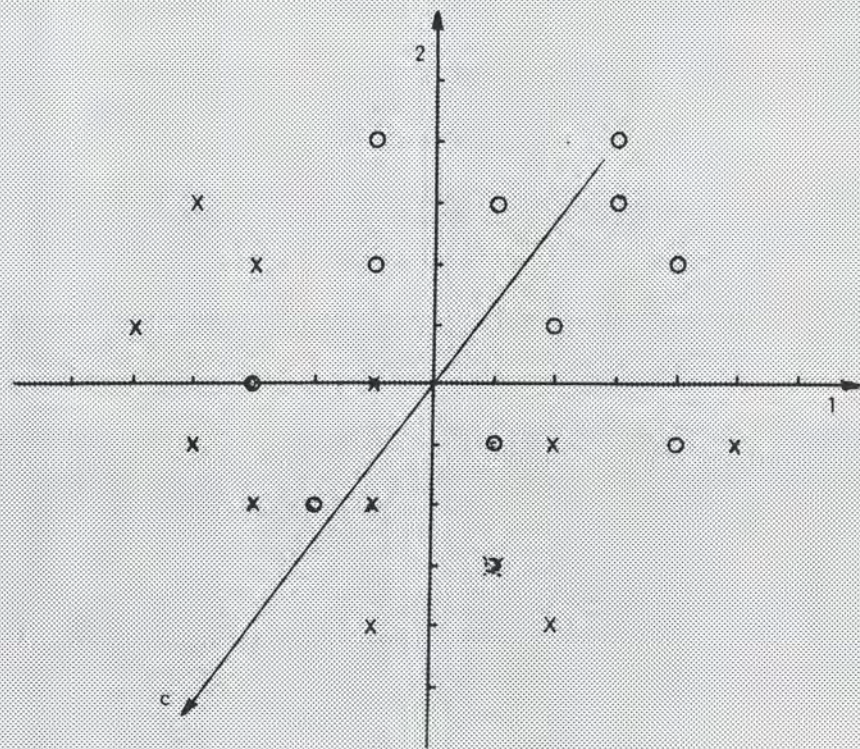


Figure 1

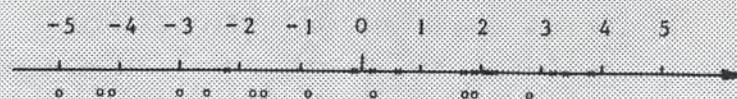


Figure 2

Légende

x = classe 1
O = classe 2

VI - Discrimination logistique

Dans notre analyse nous allons nous contenter tout comme précédemment de travailler avec deux groupes ou sous-populations.
L'objectif de l'analyse discriminante logistique est d'établir une relation entre les probabilités à posteriori et la fonction discriminante.

A) Règle de classement

Envisageons le classement d'un individu dans deux groupes Π_1 et Π_2 sur la base d'un vecteur d'observations $x' = (x_1, x_2, \dots, x_p)$. Soient p_1 et p_2 les probabilités à priori des groupes Π_1 et Π_2 ($p_1 + p_2 = 1$) et désignons par $f_1(x)$ et $f_2(x)$ les densités de probabilité de x dans chacun des deux groupes. En application du théorème de Bayes, la probabilité d'appartenir au groupe Π_2 conditionnelle au vecteur x , appelée aussi probabilité à posteriori de Π_2 , est donnée par la relation

$$\text{pr}(\Pi_2|x) = \frac{p_2 f_2(x)}{p_1 f_1(x) + p_2 f_2(x)}$$
$$\text{pr}(\Pi_2|x) = \frac{1}{1 + \frac{p_1}{p_2} \lambda(x)}; \lambda(x) = \frac{f_1(x)}{f_2(x)}$$

bien entendu,

$$\text{pr}(\Pi_1|x) = 1 - \text{pr}(\Pi_2|x)$$

La règle de classement optimale, c'est-à-dire celle qui minimise la probabilité totale de classement erroné, ou, d'une manière équivalente, qui maximise la probabilité totale de classement correct, consiste à classer l'individu dans le groupe Π_2 si

$$\text{pr}(\Pi_2|x) \geq \text{pr}(\Pi_1|x)$$

c'est-à-dire si $\lambda(x) \leq \frac{p_2}{p_1}$, et dans le groupe Π_1 sinon.

B) Détermination de la forme relationnelle entre les probabilités à posteriori et le rapport de vraisemblance

La détermination de cette forme va se faire d'abord en considérant un seul test puis ensuite plusieurs tests en même temps.

Considérons une population à laquelle sera appliquée un examen clinique.

Dans cette population figure des sujets malades M^+ (sujets atteints par la maladie que l'examen doit permettre de discriminer) et des sujets non malades M^- .

Les résultats de l'examen clinique étudiés sont résumés par une variable à deux classes c'est-à-dire la classe des sujets sains et la classe des sujets malades.

B.1) Cas d'un test unique

Cas d'un test binaire

Dans le cas d'une variable binaire (test binaire), nous dirons d'un sujet ayant donné une réponse positif à l'examen clinique qu'il a le signe positif S^+ sinon on dit qu'il est de signe négatif S^- .

Soit l'ensemble de la population repartie comme suit en tenant compte de la maladie et du résultat des tests.

MALADIE		
	M^+	M^-
S^+	P_1	P_2
S^-	P_3	P_4
	$P(M^+)$	$P(M^-)$

La prévalence P de la maladie dans cette population est égal à la probabilité pour qu'un sujet pris au hasard soit réellement atteint de cette maladie. P est aussi appelé probabilité à priori.

$$P = P(M^+) = P_1 + P_3.$$

La sensibilité Se d'une maladie est la probabilité pour un sujet malade de présenter un résultat positif c'est-à-dire la proportion de sujets malades ayant un test positif sur l'ensemble des sujets malades.

$$Se = P(S^+|M^+) = \frac{P_1}{P_1 + P_3}$$

Une sensibilité de 100% signifie que tous les sujets malades ont un test positif. En deçà de ce maximum, la sensibilité traduit l'importance relative des vrais positifs et des faux négatifs: une sensibilité forte signifie que le test donne beaucoup de vrais positifs et peu de faux positifs (la probabilité pour un sujet malade d'avoir un test positif est grand).

La Spécificité S_p d'un examen est la probabilité pour un sujet non malade de présenter un résultat négatif c'est-à-dire la proportion de sujets non malades ayant un test négatif sur l'ensemble des sujets non malades.

$$S_p = P(S^- | M^-) = \frac{P_4}{P_4 + P_2}$$

Un spécificité de 100% signifie que tous les sujets non malades ont un test négatif. En deçà de ce maximum, la spécificité traduit l'importance relative des vrais négatifs et des faux positifs: une spécificité forte signifie que le test donne beaucoup de vrais négatifs et peu de faux positifs (la probabilité qu'un sujet non malade pris au hasard d'être négatif est grand)

Rapport de vraisemblance

1) Pour un test positif

Ce rapport de vraisemblance est le rapport de la sensibilité et de la non-spécificité. Ce rapport permet de mesurer l'importance relative des vrais positifs par rapport aux faux positifs

$$L(S^+) = \frac{Se}{1 - S_p}$$

2) Pour un test négatif

Ce rapport de vraisemblance est le rapport de la spécificité et de la non sensibilité .

Ce rapport permet de mesurer l'importance relative des vrais négatifs par rapport aux faux négatifs

$$L(S^-) = \frac{S_p}{1 - Se}$$

La valeur prédictive d'un sujet appelée aussi probabilité à posteriori consiste à déterminer la probabilité d'un sujet d'être malade (respectivement non malade) sachant qu'on connaît avec certitude que le résultat du test qu'il a effectué est positif (respectivement négatif). Ces deux probabilité sont appelées respectivement valeur prédictive positive et valeur prédictive négative. Leur évaluation initiale est réalisée sur des sujets dont on connaît le statut de malade ou de sain.

Valeur prédictive positive est la probabilité pour un sujet présentant un résultat positif d'être malade c'est-à-dire la proportion de tests positifs chez les sujets malades sur l'ensemble des tests positifs.

$$P(M^+ | S^+) = \frac{P_1}{P_1 + P_2}$$

Cette valeur prédictive positive est obtenue par l'application du théorème de Bayes. Il nous est tout de même possible d'établir une relation directe la valeur prédictive positive (probabilité à posteriori) et le rapport de vraisemblance

$$\begin{aligned}
 P(M^+|S^+) &= \frac{P(M^+, S^+)}{P(S^+)} \\
 &= \frac{P(S^+|M^+).P(M^+)}{P(S^+|M^+).P(M^+) + P(S^+|M^-).P(M^-)} \\
 &= \frac{P(S^+|M^+).P(M^+)}{P(S^+|M^+).P(M^+) + P(S^+|M^-).(1 - P(M^+))} \\
 &= \frac{P(S^+|M^+).P}{P(S^+|M^+).P + P(S^+|M^-).(1 - P)} \\
 &= \frac{P.Se}{P.Se + (1 - P).(1 - Sp)} \\
 &= \frac{P.L(S^+)}{P.L(S^+) + (1 - P)}
 \end{aligned}$$

La valeur prédictive négative est la probabilité pour un sujet présentant un résultat négatif d'être en bonne santé c'est-à-dire la proportion de test négatif chez les sujets non malades sur l'ensemble des tests négatifs.

$$P(M^-|S^-) = \frac{P_4}{P_3 + P_4}$$

Cette valeur prédictive négative est aussi obtenue par l'application du théorème de Bayes. En appliquant la même démarche que précédemment on arrive à établir une relation directe entre la valeur prédictive négative (probabilité à posteriori) et le rapport de vraisemblance:

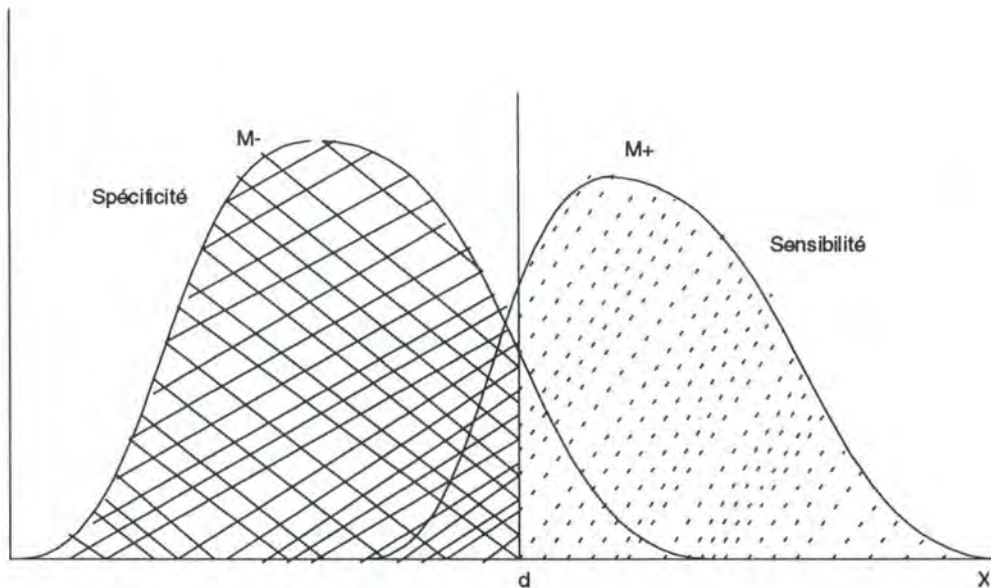
$$\begin{aligned}
 P(M^-|S^-) &= \frac{(1 - P).Sp}{P.(1 - Se) + (1 - P).(1 - Sp)} \\
 &= \frac{(1 - P).L(S^-)}{P + (1 - P).L(S^-)}
 \end{aligned}$$

Remarque: Dans cadre d'un test binaire, on constate qu'il existe une relation directe entre la probabilité à posteriori et le rapport de vraisemblance.

Cas d'un test quantitatif Gaussien

Spécificité et sensibilité

Dans le cas d'une variable quantitative gaussienne on détermine la courbe de densité de probabilité de la spécificité et la courbe de densité de probabilité de la sensibilité (voir figure ci-dessous)



Comme on le sait déjà, un test laboratoire est évalué en terme de sa capacité à discriminer entre une population en bonne santé M^- et une population malade M^+ . Pour positionner donc un patient à la suite d'un résultat de son examen, on déduit une variable binaire y et un point de décision d diagnostic tel que

$$\begin{array}{l} | S^- \text{ ou } y = 0 \text{ (examen négatif) quand } x \leq d \\ | S^+ \text{ ou } y = 1 \text{ (examen positif) quand } x > d \end{array}$$

Ce qui nous ramène à l'analyse effectuée avec le test binaire précédemment, à la différence que le positionnement d'un patient qui a effectué un examen clinique dépend du choix du point de décision diagnostic d .

Ainsi donc la sensibilité et la spécificité dépendent aussi du choix du point de décision diagnostic.

La spécificité étant définie comme étant la proportion de personnes en bonne santé pour qui on a $y = 0$ ($x \leq d$). Ceux-ci sont de "vrais négatifs". Leur proportion est l'aire de la distribution M^- à gauche du point de décision d .

La sensibilité d'un test est défini pour être la proportion de la population malade pour qui $y = 1$ ($x > d$). Ce sont de "vrais positifs" et leur proportion est l'aire de distribution M^+ à droite de l'ordonnée $x = d$.

Ainsi donc la sensibilité et la spécificité dépendent du choix du point de décision. On peut les exprimer comme suite

$$\text{Spécificité: } S_p = P(y = 0|M^-, d) = S_p(d)$$

$$\text{Sensibilité: } Se = P(y = 1|M^+, d) = Se(d)$$

En déplaçant le point de décision d de la droite vers la gauche, la spécificité décroît alors que la sensibilité croît. Inversement, le déplacement ce point de la gauche vers la droite entraîne une croissance de la spécificité et une décroissance de la sensibilité.

Rapport de vraisemblance

1) Pour le signe positif

Ce rapport permet de mesurer l'importance relative des vrais positifs par rapport aux faux positifs.

$$L(y = 1) = L(1) = \frac{P(y = 1|M^+)}{P(y = 1|M^-)} = \frac{P(y = 1|M^+)}{1 - P(y = 0|M^-)} = \frac{Se(d)}{1 - S_p(d)}$$

2) Pour le signe négatif

Ce rapport permet de mesurer l'importance relative des vraies négatifs par rapport aux faux négatifs.

$$L(y = 0) = L(0) = \frac{P(y = 0|M^-)}{P(y = 0|M^+)} = \frac{P(y = 0|M^-)}{1 - P(y = 1|M^+)} = \frac{S_p(d)}{1 - Se(d)}$$

La valeur prédictive

Valeur prédictive positive:

$$\begin{aligned} P(M^+|y = 1) &= \frac{P(y = 1, M^+)}{P(y = 1)} = \frac{P(M^+) \cdot P(y = 1|M^+)}{P(M^+) \cdot P(y = 1|M^+) + P(M^-) \cdot P(y = 1|M^-)} \\ &= \frac{P \cdot P(y = 1|M^+)}{P \cdot P(y = 1|M^+) + (1 - P) \cdot P(y = 1|M^-)} \\ &= \frac{P \cdot Se(d)}{P \cdot Se(d) + (1 - P) \cdot (1 - S_p(d))} \\ &= \frac{P \cdot L(1)}{P \cdot L(1) + (1 - P)} \end{aligned}$$

où $P = P(M^+)$ est la prévalence ou probabilité à priori.

Valeur prédictive négative

$$\begin{aligned}
 P(M^- | y = 0) &= \frac{P(y = 0, M^-)}{P(y = 0)} = \frac{P(M^-) \cdot P(y = 0 | M^-)}{P(M^+) \cdot P(y = 0 | M^+) + P(M^-) \cdot P(y = 0 | M^-)} \\
 &= \frac{(1 - P) \cdot S_p(d)}{(1 - P) \cdot S_p(d) + P \cdot (1 - Se(d))} \\
 &= \frac{1 - P}{P \cdot L(0) + (1 - P)}
 \end{aligned}$$

Remarque:

On constate que la relation directe précédemment établie entre la valeur prédictive (probabilité à posteriori) et le rapport de vraisemblance est une fois encore vérifiée.

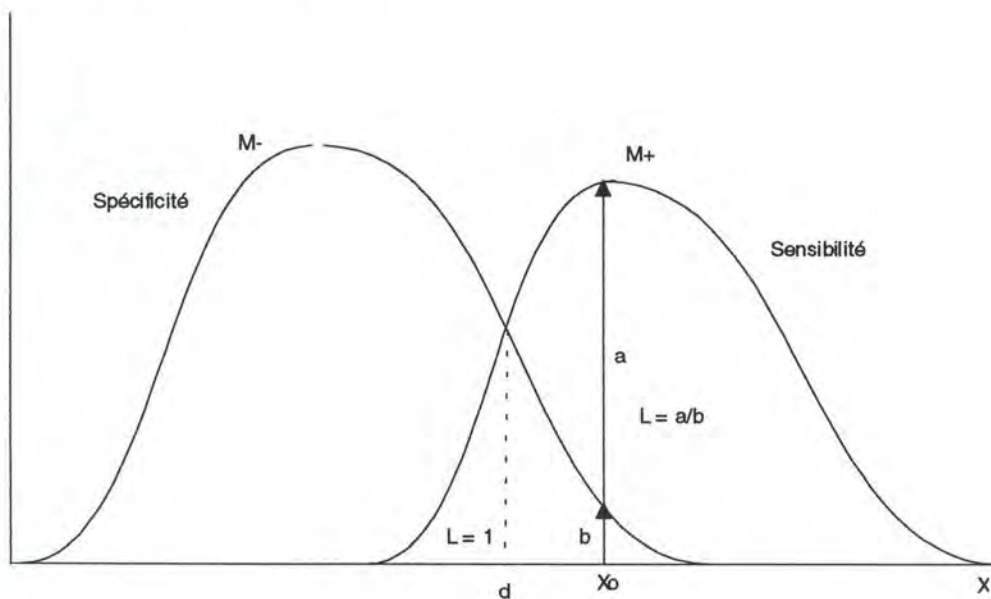
Nous savons par hypothèses que les valeurs de tests prises par les individus sains aussi bien que malades suivent la loi normale et que

- | S- ou $y = 0$ (examen négatif) quand $x \leq d$
- | S+ ou $y = 1$ (examen positif) quand $x > d$

Le fait de déterminer le rapport de vraisemblance associé à y revient à déterminer le rapport de vraisemblance associé à x .

Ainsi $L(x) = L(y)$

Au vue de la figure ci-dessous



on remarque que le rapport de vraisemblance associé à la valeur observée x_0 de test laboratoire X est :

$$L(x_0) = \frac{a}{b} = \frac{P(x_0|M^+)}{P(x_0|M^-)}$$

On peut constater que le point de décision d correspond un rapport de vraisemblance égal à 1: $L(d) = 1$;

Ainsi lorsque $L(x_0) > 1$ cela traduit que $a > b$ c'est-à-dire que la probabilité que le patient qui a la valeur x_0 appartient au groupe des malades paraît plus vraisemblable.

Ainsi lorsque $L(x_0) = 1$ cela traduit le fait que $a = b$ c'est-à-dire la probabilité que le patient qui a la valeur x_0 appartient au groupe des sains est identique à la probabilité que ce même patient ayant la valeur x_0 appartient au groupe des malades. Cela devient difficile d'affecter le patient à un groupe donnée.

Dans ce cas la décision est laissée au médecin qui connaît mieux son patient et qui a sûrement à sa disposition des atouts que le logiciel qui sera mise en place ne peut pas incorporer comme par exemple l'expérience du médecin en la matière

RESUME

On peut constater que si on appelle M^+ la maladie pour laquelle on réalise l'analyse clinique et x l'examen clinique pratiqué sur un patient quelconque, la valeur prédictive(probabilité à posteriori) d'un patient quelconque qui a pour but de prédire sur base des examens cliniques si un patient est malade ou pas. En d'autre terme cela détermine la probabilité qu'un patient soit malade sachant qu'il a subi un test x, s'écrit toujours quel que soit le type test de la façon suivante:

$$P(M^+|X) = \frac{P.L(X)}{PL(X) + (1 - P)}$$

où P est appelé la prévalence(probabilité à priori) c'est-à-dire la probabilité qu'un patient pris au hasard dans la population des patient soit malade. Cela correspond à

$$P = P(M^+) = \frac{n^+}{n^+ + n^-} = \frac{n^+}{n}$$

avec n représentant la population totale des patients subdivisée en deux sous populations de patients effectivement malades n^+ et de patients non malades n^- .

$L(x)$ représente le rapport de vraisemblance. Ce rapport est toujours positif où nul. Cela correspond par définition à

$$L(X) = \frac{P(X|M^+)}{P(X|M^-)}$$

Cette relation établie entre la probabilité à posteriori et le rapport de vraisemblance valable dans le cas d'un seul test est-elle valable dans le cas de plusieurs tests réalisés sur les patients dans notre analyse?

B.2) Cas de plusieurs tests

Dans ce cas il se peut que les différents tests soient réalisés par le médecin sur l'ensemble des patients ne soient constitués que de tests binaires, soit de quantitatifs où soit d'un mélange des deux types de tests.

Dans ce cas l'ensemble des tests réalisés sur un patient donné fait l'objet d'une interprétation simultanée.

Soit X , ce profil constitué de plusieurs tests indépendants x_i : $X = (x_1, x_2, x_3, \dots, x_n)$.
Comme les tests x_i sont indépendants alors :

$$P(X | M^+) = P(x_1 | M^+) P(x_2 | M^+) P(x_3 | M^+) \dots P(x_n | M^+)$$

De façon similaire on a

$$P(X | M^-) = P(x_1 | M^-) P(x_2 | M^-) P(x_3 | M^-) \dots P(x_n | M^-)$$

Comme par définition

$$L(x_i) = \frac{P(x_i | M^+)}{P(x_i | M^-)} \quad \text{pour } i = 1, 2, \dots, n$$

alors on obtient

$$L(X) = L(x_1) L(x_2) \dots L(x_n)$$

Ainsi donc, le rapport de vraisemblance d'un profil X de tests indépendants est égal au produit de chacun des tests cet ensemble.

Si on venait à supposer que le profil X était formé d'un seul test on aurait

$$P(M^+ | X) = \frac{P \cdot L(X)}{P \cdot L(X) + (1 - P)}$$

En remplaçant $L(X)$ par sa valeur on aurait

$$P(M^+ | X) = \frac{P \prod_{i=1}^n L(x_i)}{P \prod_{i=1}^n L(x_i) + (1 - p)}$$

Soit X un profil constitué de deux tests x_1 et x_2 dépendants.

Alors la probabilité $P(X | M^+)$ peut se décomposer sous forme de probabilité conditionnelle dont le résultat est

$$P(X | M^+) = P(x_1 | x_2 | M^+).P(x_2 | M^+)$$

De façon similaire on obtient pour la probabilité $P(X | M^-)$ le résultat suivant

$$P(X | M^-) = P(x_1 | x_2 | M^-).P(x_2 | M^-)$$

Le rapport de vraisemblance $L(X)$ se dégageant du rapport de $P(X | M^+)$ et $P(X | M^-)$ nous donne le résultat suivant

$$\begin{aligned} L(X) &= \frac{P(x_1|x_2|M^+).P(x_2|M^+)}{P(x_1|x_2|M^-).P(x_2|M^-)} \\ &= \frac{P(x_1|x_2|M^+)}{P(x_1|x_2|M^-)} \times \frac{P(x_2|M^+)}{P(x_2|M^-)} \\ &= L(x_1 | x_2).L(x_2) \end{aligned}$$

Ainsi donc le rapport de vraisemblance obéit à des règles similaires de celles des probabilités conditionnelles.

En généralisant le profil de deux tests à un profil de n tests on obtient le résultat suivant

$$L(X) = L(x_1 | x_2, x_3, \dots, x_i, \dots, x_n).L(x_2 | x_3, \dots, x_i, \dots, x_n) \dots L(x_i | x_{i+1}, \dots, x_n) \dots L(x_n)$$

En remplaçant $L(X)$ par sa valeur dans la relation liant la probabilité à posteriori et le rapport de vraisemblance on obtient le résultat suivant

$$P(M^+|X) = \frac{P.L(x_1|x_2x_3\dots x_n).L(x_2|x_3x_4\dots x_n).....L(x_i|x_{i+1}\dots x_n).....L(x_n)}{P.L(x_1|x_2x_3\dots x_n).L(x_2|x_3x_4\dots x_n).....L(x_i|x_{i+1}\dots x_n).....L(x_n) + (1 - P)}$$

Force est de constater que la relation qui lie la probabilité à posteriori $P(M^+|X)$ et le rapport de vraisemblance $L(X)$

$$P(M^+|X) = \frac{P.L(X)}{P.L(X) + (1 - P)}$$

est valable quelque soit les types de tests et le nombre de tests. L'objectif est d'automatiser le calcul de cette probabilité $P(M^+|X)$. Cela n'est possible que lorsque ce rapport de vraisemblance est facile à déterminer.

Par définition, ce rapport de vraisemblance $L(X)$ égal à

$$L(X) = \frac{P(X|M^+)}{P(X|M^-)}$$

devient difficile à déterminer lorsque le nombre de tests devient nombreux et que ces tests sont indépendants ou non.

Le problème sera de trouver une forme fonctionnelle simple du rapport de vraisemblance qui ne tient pas compte du type de distribution et du type test.

C) Détermination de la forme du rapport de vraisemblance

Pour déterminer cette forme fonctionnelle du rapport de vraisemblance, nous allons procéder par étape en cherchant d'abord la forme fonctionnelle de ce rapport selon qu'on a affaire à un seul test et puis ensuite lorsque nous avons affaire à plusieurs tests. Dans ce dernier cas que devient cette forme fonctionnelle lorsque les tests sont indépendants ou non.

C.1) Cas de test unique

Test Gaussien

$$P(X/M^+) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{X - \bar{X}_{M^+}}{\sigma}\right)^2\right)$$

$$P(X/M^-) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{X - \bar{X}_{M^-}}{\sigma}\right)^2\right)$$

$$\begin{aligned} L(X) &= \frac{P(X/M^+)}{P(X/M^-)} = \frac{\exp\left(-\frac{1}{2}\left(\frac{X - \bar{X}_{M^+}}{\sigma}\right)^2\right)}{\exp\left(-\frac{1}{2}\left(\frac{X - \bar{X}_{M^-}}{\sigma}\right)^2\right)} \\ &= \exp\left(-\frac{1}{2}\left(\frac{X - \bar{X}_{M^+}}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{X - \bar{X}_{M^-}}{\sigma}\right)^2\right) \\ &= \exp\left(\frac{-1}{2\sigma^2}(\bar{X}_{M^+}^2 - \bar{X}_{M^-}^2) + \frac{1}{\sigma^2}(\bar{X}_{M^+} - \bar{X}_{M^-})X\right) \\ &= \exp(\alpha_0 + \alpha_1 X) \end{aligned}$$

On remarque le rapport de vraisemblance débouche sur une fonction qui est une forme exponentielle: $\exp(\alpha_0 + \alpha_1 X)$

Test binaire

Dans le cas d'un test unique on a trouvé que pour :

$$X = 1 \quad L(X) = \frac{1 - Se}{S_p}$$

$$X = 0 \quad L(X) = \frac{Se}{1 - S_p}$$

Cherchons à déterminer la fonction de vraisemblance $L(X)$. Donc il s'agit de déterminer une fonction $L(X)$ telle que

$$L(0) = \frac{Se}{1 - S_p}$$

et

$$L(1) = \frac{1 - Se}{S_p}$$

Supposons que la fonction $L(X)$ a une forme exponentielle comme suite: $L(X) = \text{Exp}(\alpha_0 + \alpha_1 X)$
Dans ce cas :

$$L(0) = \text{Exp}(\alpha_0) = \frac{1 - Se}{S_p} \Rightarrow \alpha_0 = \text{Log}\left(\frac{1 - Se}{S_p}\right)$$

$$L(1) = \text{Exp}(\alpha_0 + \alpha_1) = \frac{Se}{1 - S_p} \Rightarrow$$

$$\alpha_1 = \text{Log}\left(\frac{Se}{1 - S_p}\right) - \text{Log}\left(\frac{1 - Se}{S_p}\right) = \text{Log}\left(\frac{Se}{1 - S_p} \times \frac{1 - Se}{S_p}\right)$$

On remarque ici que la fonction de vraisemblance épouse toujours une forme exponentielle comme précédemment. Les coefficients de cette fonction dépendent de la moyenne de la population de sujets malades et de la variance de chacune des populations supposées identiques pour ces deux populations.

C.2) Cas de tests multiples

On a vu dans l'analyse précédente qu'en supposant les tests indépendants entre eux alors le rapport de vraisemblance s'écrit

$$L(X) = L(x_1) \cdot L(x_2) \cdot L(x_3) \dots L(x_n)$$

Comme chaque rapport de vraisemblance $L(X_i)$ est de la forme

$$L(X_i) = \text{Exp}(\alpha_{0i} + \alpha_i X_i) \quad i = 1, 2, \dots, n$$

Alors le rapport de vraisemblance $L(X)$ de l'ensemble des tests indépendants sera

$$\begin{aligned} L(X) &= \prod_{i=1}^n L(X_i) = \prod_{i=1}^n \text{Exp}(\alpha_{0i} + \alpha_i X_i) \\ &= \text{Exp}\left(\sum_{i=1}^n (\alpha_{0i} + \alpha_i X_i)\right) = \text{Exp}(\alpha_0 + \alpha^T X) \end{aligned}$$

Selon le professeur Adelin Albert (cf. On the Use and Computation of Likelihood Ratios in Clinical Chemistry CLINICAL CHEMISTRY, Vol. 28, N° 5, 1982) La forme exponentielle de la fonction de vraisemblance reste toujours valable même lorsque les tests ne sont pas indépendants.

$$L(X) = \text{Exp}\left(\sum_{i=1}^n (\alpha_{0i} + \alpha_i X_i)\right) = \text{Exp}(\alpha_0 + \alpha^T X)$$

Avec

$$\alpha_0 = \sum_i \alpha_{0i} \quad \text{et} \quad \alpha^T = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n)$$

D) Détermination de la forme logistique

D'après l'analyse précédemment réalisée, on a montré que la valeur de prévalence (probabilité à posteriori) était de la forme suivante

$$P(M^+|X) = \frac{P \cdot L(X)}{P \cdot L(X) + (1-P)} = \frac{1}{1 + \frac{1-P}{P} \frac{1}{L(X)}} \quad (D.1)$$

et que le rapport de vraisemblance $L(X)$ avait une forme exponentielle

$$L(X) = \text{Exp}(\alpha_0 + \alpha^T X)$$

où

$\alpha_0 + \alpha^T X$ est appelé fonction linéaire discriminante

En remplaçant la valeur de $L(X)$ dans l'équation (D.1) on obtient alors la forme de la valeur de prévalence (probabilité à priori) suivante

$$\begin{aligned} P(M^+|X) &= \frac{1}{1 + \frac{1-P}{P} \text{Exp}(-\alpha_0 - \alpha^T X)} \\ &= \frac{1}{1 + \text{Exp}(\beta_0 + \beta^T X)} \end{aligned}$$

Avec

$$\beta_0 = \text{Log}\left(\frac{1-P}{P}\right) - \alpha_0 \quad \text{et} \quad \beta^T = -\alpha^T$$

Ainsi la fonction

$$P(M^+ | X) = \frac{1}{1 + \text{Exp}(\beta_0 + \beta^T X)}$$

est appelée fonction logistique discriminante.

La probabilité à posteriori $P(M|X)$ donne le résultat suivant

$$P(M^- | X) = 1 - P(M^+ | X)$$

$$P(M^- | X) = \frac{\text{Exp}(\beta_0 + \beta^T X)}{1 + \text{Exp}(\beta_0 + \beta^T X)}$$

E) Estimation des coefficients de la fonction logistique

En pratique, le problème se pose d'estimer les coefficients de la fonction logistique discriminante $\beta_0 + \beta^T$. Pour clarifier les choses, il convient de faire une distinction suivant que l'échantillon d'effectif n dont dispose est extrait du mélange des deux groupes M^+ (l'ensemble des patient malades) et M^- (l'ensemble des patient non malades) ou séparément de chacun des groupes. En effet si n_1 et n_2 désignent respectivement des échantillons de M^+ et M^- , dans le premier cas n_1 et n_2 sont aléatoires et dans le second ils sont fixés. D'un point de vue statistique, cette distinction essentielle.

E.1) Echantillon composé du mélange des deux groupes

Envisageons d'abord le problème d'estimation lorsque l'utilisateur dispose d'un échantillon de n observations extrait du mélange de la population des sujets malades M^+ et des sujets non malades M^- . A cet effet, on introduit une variable Z telle que

$Z = 0$, si l'individu provient de M^-
 $Z = 1$, si l'individu provient de M^+

L'échantillon peut alors s'écrire
 $\{(Z_i, X_i), i = 1, \dots, n\}$

où $X_i = (X_{i1}, X_{i2}, \dots, X_{ik})$ désigne le vecteur d'observation du $i^{\text{ème}}$ individu.
la probabilité lié à l'échantillon est

$$\begin{aligned} Q(\beta_0, \beta) &= \pi \prod_{i=1}^n [P(M^-|X)]^{1-z_i} \cdot [P(M^+|X)]^{z_i} \\ &= \prod_{i=1}^n \left(\frac{\text{Exp}(\beta_0 + \beta^T X_i)}{1 + \text{Exp}(\beta_0 + \beta^T X_i)} \right)^{1-z_i} \cdot \left(\frac{1}{1 + \text{Exp}(\beta_0 + \beta^T X_i)} \right)^{z_i} \end{aligned}$$

la maximisation de $Q(\beta_0, \beta)$ nous permettra d'obtenir les estimateurs de maximum de vraisemblance de β_0 et de β^T c'est-à-dire les coefficients de la fonction discriminante. Pour résoudre le problème, il faut maximiser $Q(\beta_0, \beta)$ ou son logarithme.

Les estimateurs du maximum de vraisemblance s'obtiennent en résolvant les équations

$$\frac{\partial \text{Log} Q(\beta_0, \beta)}{\partial \beta_s} = 0 \quad \text{où } s = 0, 1, \dots, k$$

qui s'écrivent aussi, après quelques développements en

$$\sum_{i \in E^2} \left(\frac{\text{Exp}(\xi_i)}{1 + \text{Exp}(\xi_i)} \right) X_{is} - \sum_{i \in E^1} \left(\frac{1}{1 + \text{Exp}(\xi_i)} \right) X_{is} \quad \text{où } s = 0, 1, \dots, k$$

En rappelant que

$$\xi_i = \beta_0 + \beta^T X_i \quad E_1 = \{i : Z_i = 0\} \text{ et } E_2 = \{i : Z_i = 1\}.$$

Ce système de $k+1$ équation à $k+1$ inconnues n'admet pas de solution analytique. Afin de déterminer le point stationnaire $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k)$, on a recours à la méthode classique de NEWTON-RAPHSON définie par le schéma itératif suivant

$$\hat{\beta}_{t+1} = \hat{\beta}_t - H^{-1}(\hat{\beta}_t) \cdot C(\hat{\beta}_t) \quad t = 0, 1, 2, \dots$$

$\hat{\beta}_t$ est l'estimateur du vecteur β à la $t^{\text{ième}}$ itération du processus.

$C(\hat{\beta}_t)$ est le vecteur des dérivées premières de $\text{Log}Q(\beta_0, \beta)$ en $\beta = \hat{\beta}_t$,

$H(\hat{\beta}_t)$ est la matrice des dérivées secondes $\frac{\partial^2 \text{Log}Q(\beta_0, \beta)}{\partial \beta_i \partial \beta_j}$ en $\hat{\beta} = \hat{\beta}_t$,

Remarques pratiques:

Ces remarques pratiques portent sur la résolution des équations de maximum de vraisemblance.

i) On constate qu'au cours du processus d'itératif la matrice H^{-1} varie très lentement en comparaison du vecteur gradient C . Il n'est donc pas nécessaire de la calculer à chaque itération, ce qui dans certain cas, permet de réaliser un gain de temps appréciable.

ii) Dans tout processus itératif, il convient de choisir un point de départ approprié. Toutefois, l'expérience montre qu'en partant de l'origine c'est-à-dire $\hat{\beta}_{t=0} = (0, 0, \dots, 0)$, il est rare que l'on ne converge pas après plusieurs itérations.

E.2) Echantillons séparés

Supposons, à présent, que l'on dispose d'échantillons extraits séparément de chacun des deux groupes M^+ et M^- .

Dans ce cas les effectifs n_1 et n_2 sont fixés et n'apportent aucune information quant à la proportion relative de chacun des groupes dans le mélange.

Cette situation est la plus fréquemment rencontrée en pratique.

Il est alors indispensable de définir soi-même les probabilités à priori P et $(1-P)$, soit qu'elles sont connues, soit qu'elles sont estimées par un autre échantillon. Dans le cas d'échantillons séparés, J.A. Anderson a pu démontrer que pour obtenir les estimateurs de maximum de vraisemblance des coefficients β à partir d'échantillons séparés, il suffit d'utiliser la méthode décrite précédemment comme si l'échantillon était extrait du mélange et de faire ensuite un ajustement du terme indépendant $\underline{\beta_0}$.

Ainsi la relation entre les estimateurs $(\hat{\beta}_0^*, \hat{\beta}^*)$ de maximum de vraisemblance des coefficients de la fonction logistique dans le cadre des échantillons séparés et les estimateurs $(\hat{\beta}_0, \hat{\beta})$ de maximum de vraisemblance des coefficients de la fonction logistique dans le cadre des échantillons issus d'un mélange est donnée par, pour

$$P(M^+|X) = \frac{1}{1 + \text{Exp}(\hat{\beta}_0^* + \hat{\beta}^{*T} X)}$$

on a comme résultat :

$$\hat{\beta}_0^* = \hat{\beta}_0 + \text{Log}\left(\frac{P}{1-P}\right)$$

$$\hat{\beta}^* = \hat{\beta}$$

F) Exemple d'application

La maladie étudiée ici porte sur l'hépatite viral qui est une maladie du foie. Pour cette maladie on a considéré deux tests (AST et ALT). Les données ci-dessous portent sur des mesures effectuées sur différents patients pour ces deux tests.

L'introduction de ces données dans la machine nous a permis d'obtenir les résultats qui suivront par la suite.

Identification des patients				Valeurs des variables tests	
Noms et prénoms	Groupe	Sexe	Date naissance	Test1: AST	Test2: ALT
Ekissi jean	1	M	180245	236	582
Adehi paul	1	M	120665	65	258
Lecoq paul	1	M	140245	59	244
Niago eloi	1	M	120667	92	120
Cheridan thomas	1	M	120546	99	352
Djebi atsain	1	M	020560	87	300
Brou firmin	1	M	120565	202	429
Brou euprasie	1	F	101267	208	539
Brou justin	1	M	120359	105	465
Koffi jeremi	1	M	050847	95	442
Chauville steph	1	F	021274	215	570
Ayissi carine	1	F	020541	101	277
Dagouri martine	1	F	171155	454	639
Daryl justine	1	F	121255	308	1243
Karim faustin	1	M	140365	82	294
Devry laurent	1	M	120755	196	760
Dubois florent	1	M	170863	141	550
Bah souleymane	1	M	120272	73	201
Lou corentin	1	M	030654	569	1354

Sylla pauline	1	F	120658	125	390
Pouky ella	1	F	051265	77	480
Togbe pierre	1	M	120566	78	495
Sheila pierrette	1	F	040232	165	330
Binard marcel	1	M	250247	214	680
Curie simone	1	F	270458	141	539
Semba justin	1	M	280463	220	990
Sylla marie	1	F	240163	127	480
Dupré paul	1	M	020455	324	936
Begré laurence	1	F	120547	149	334
Kepre éloi	1	M	061272	270	465
Dela élie	1	F	210656	178	660
Deliège jeanot	1	M	120345	202	693
Demer guy	1	M	250145	56	158
Evans julien	1	M	250470	129	704
Brown james	1	M	200153	682	1595
Lamar kelly	1	F	190441	290	987
Abort victor	1	M	210836	61	327
Newman victoria	1	F	031245	289	693
Desel lola	1	F	150765	133	470
Ekissi chantal	1	F	151262	170	620
Adehi thomas	1	M	100670	550	1034
Evans cheryl	1	F	140361	313	414
Diagne papis	1	M	021265	166	239
Amon louis	1	M	150264	340	840
Binard véronica	1	F	311242	660	1287
Deville vincent	1	M	120547	321	1254
Devilliers syryl	1	M	020658	401	492
Lecharlier éliane	1	F	120564	665	970
Shell ella	1	F	280763	242	528
Chipi marcel	1	M	120565	339	672
Kablan eloi	1	M	051241	372	910
Deville yvonne	1	F	120565	165	584
Kablan coulé	1	M	120545	385	610
Daryl célestine	1	F	120654	152	270
Karim danielle	1	F	120545	605	1199
Bony hélène	1	F	150236	247	453
Niango élalie	2	F	021263	53	123
Diallo aminata	1	F	021255	76	133
Dindja sali	2	F	120565	31	63
Karim salimanta	2	F	120564	32	56
Karim cécile	2	F	120563	50	59
Brou célestin	2	M	120475	56	72
Lecoq louis	2	M	120765	39	87
Dagou béatrice	2	F	120475	46	95

Delpot paul	2	M	140365	29	57
Durango pierre	2	M	140263	40	50
Durango sylvie	2	F	021264	29	44
Durant eveline	2	F	121289	77	132
Dupont liliane	2	F	120565	28	44
Koffi julien	2	M	120245	28	68
Lian édouard	2	M	120545	34	58
Lian sylvie	2	F	120545	24	42
Dubois laurence	2	F	101251	34	61
Chavanne marc	2	M	120541	119	345
Luppi sylvain	2	M	120473	26	41
Piret sophie	2	F	120365	39	84
Piquet laura	2	F	120656	30	70
Filou gauthier	2	M	140256	35	81
Dechamp lisa	2	F	120545	29	41
Leheureux marc	2	M	210547	31	38
Finidi laureine	2	F	240381	31	56
Herpelin tina	2	F	300564	23	43
Ducro paul	2	M	120261	38	40
Durant lise	2	F	121258	22	42
Deville désiré	2	M	121256	26	57
Lemaire yvonne	2	F	120545	26	42
Dechamp marco	2	M	120545	35	72
Ducro claudine	2	F	120645	42	53
Marloi cécile	2	F	020545	41	47
Flecheur florentine	2	F	120547	64	260
Filou céline	2	F	021155	212	660
Sié marcellin	2	M	050247	83	101
Bah lise	2	F	120265	64	185
Lou corentine	2	F	020445	121	159
Piret émile	2	M	120468	29	42
sika lazare	2	M	020545	129	163
Sylla rachel	2	F	100265	76	109
Piquet laurent	2	M	140362	135	118
Séka clément	1	M	140654	207	83
Koné abou	2	M	110550	628	23
piquet sophie	1	F	120255	802	25
Kassy paul	2	M	120365	1254	321
Baka soul	2	M	120554	100	2000

Choix d'analyse

Type d'échantillon: 0 (échantillon provenant d'un mélange)

Nombre de tests: 2

Maladie

Code: # HEPATITE

Nom: HEPATITE VIRAL

Nom des variables tests

Test1: AST(aspartate aminotransferase)

Test2: ALT (alanine aminotransferase)

Statistiques sur les patients

Nbr. de malades: 44

Nbr. de sain : 59

Nbr. total de patients: 103

Maladie: HEPATITE VIRAL

Probabilité des malades: 0.427

Probabilité des sains: 0.573

Statistiques des variables tests étudiés

Ecart types	Moyennes	Variables tests
203.261	180.515	AST
394.179	398.835	ALT

Matrice des valeurs moyennes

		Patients	
		Sains	Malades
Variables tests	Test1	245.339	587.729
	Test2	93.591	145.545

Matrice des écart types

		Patients	
		Sains	Malades
Variables	Test1	179.493	344.112
tests	Test2	202.605	306.991

Coéfficients de la fonction de vraisemblance

Test1(AST): -0.208738

Test2(ALT): 0.163688

Constante : 1.250000

Allocation abregée

			Patients	
			Sains	Malades
Sains	56	4		
Malades	3	40		
Total	59	44		

Allocation détaillée

Identification des patients	Répartition des patients		Probabilités	
	Gr. réel	Gr. alloué	Gr. sain=1	Gr. malade=2
Noms et prenom				
Ekissi jean	1	1	1.000000	0.000000
Adehi paul	1	1	1.000000	0.000000
Lecoq paul	1	1	1.000000	0.000000
Niogo eloi	1	1	0.844054	0.155946
Cheridan thomas	1	1	1.000000	0.000000
Djebi atsain	1	1	1.000000	0.000000
Brou firmin	1	1	1.000000	0.000000
Brou euprasie	1	1	1.000000	0.000000
Brou justin	1	1	1.000000	0.000000
Koffi jeremi	1	1	1.000000	0.000000
Chauville steph	1	1	1.000000	0.000000
Ayissi carine	1	1	1.000000	0.000000
Dagouri martine	1	1	0.999985	0.000015
Daryl justine	1	1	1.000000	0.000000
Karim faustin	1	1	1.000000	0.000000

Devry laurent	1	1	1.000000	0.000000
Dubois florent	1	1	1.000000	0.000000
Bah souleymane	1	1	1.000000	0.000000
Lou corentin	1	1	1.000000	0.000000
Sylla pauline	1	1	1.000000	0.000000
Pouky ella	1	1	1.000000	0.000000
Togbe pierre	1	1	1.000000	0.000000
Sheila pierrette	1	1	1.000000	0.000000
Binard marcel	1	1	1.000000	0.000000
Curie simone	1	1	1.000000	0.000000
Semba justin	1	1	1.000000	0.000000
Sylla marie	1	1	1.000000	0.000000
Dupré paul	1	1	1.000000	0.000000
Begré laurence	1	1	1.000000	0.000000
Kepré éloi	1	1	1.000000	0.000000
Dela élie	1	1	1.000000	0.000000
Deliège jeanot	1	1	1.000000	0.000000
Demer guy	1	1	1.000000	0.000000
Evans julien	1	1	1.000000	0.000000
Brown james	1	1	1.000000	0.000000
Lamar kelly	1	1	1.000000	0.000000
Abort victor	1	1	1.000000	0.000000
Newman victoria	1	1	1.000000	0.000000
Desel lola	1	1	1.000000	0.000000
Ekissi chantal	1	1	1.000000	0.000000
Adehi thomas	1	1	1.000000	0.000000
Evans cheryl	1	1	0.975445	0.024555
Diagne papis	1	1	0.996734	0.003266
Amon louis	1	1	1.000000	0.000000
Binard véronica	1	1	1.000000	0.000000
Deville vincent	1	1	1.000000	0.000000
Devilliers syryl	1	2	0.127944	0.872056
Lecharlier éliane	1	1	1.000000	0.000000
Shell ella	1	1	1.000000	0.000000
Chipi marcel	1	1	1.000000	0.000000
Kablan eloi	1	1	1.000000	0.000000
Deville yvonne	1	1	1.000000	0.000000
Kablan eoulé	1	1	1.000000	0.000000
Daryl célestine	1	1	0.999999	0.000001
Karim danielle	1	1	1.000000	0.000000
Bony hélène	1	1	1.000000	0.000000
Niango élalie	2	2	0.000033	0.999967
Diallo aminata	1	1	0.999221	0.000779
Dindja sali	2	2	0.006111	0.993889
Karim salimanta	2	2	0.023273	0.976727

Karim cécile	2	2	0.384447	0.615553
Brou célestin	2	2	0.206487	0.793513
Lecoq louis	2	2	0.000642	0.999358
Dagou béatrice	2	2	0.000747	0.999253
Delpot paul	2	2	0.010699	0.989301
Durango pierre	2	2	0.252580	0.747420
Durango sylvie	2	2	0.083258	0.916742
Durant eveline	2	2	0.001130	0.998870
Dupont liliane	2	2	0.068650	0.931350
Koffi julien	2	2	0.001448	0.998552
Lian édouard	2	2	0.025411	0.974589
Lian sylvie	2	2	0.042485	0.957515
Dubois laurence	2	2	0.015706	0.984294
Chavanne marc	2	2	0.000000	1.000000
Luppi sylvain	2	2	0.073507	0.926493
Piret sophie	2	2	0.001049	0.998951
Piquet laura	2	2	0.001584	0.998416
Filou gauthier	2	2	0.000744	0.999256
Dechamp lisa	2	2	0.129226	0.870774
Leheureux marc	2	2	0.269083	0.730917
Finidi laureine	2	2	0.018971	0.981029
Herpelin tina	2	2	0.029666	0.970334
Ducro paul	2	1	0.533576	0.466424
Durant lise	2	2	0.028397	0.971603
Deville désiré	2	2	0.005749	0.994251
Lemaire yvonne	2	2	0.063108	0.936892
Dechamp marco	2	2	0.003237	0.996763
Ducro claudine	2	2	0.238942	0.761058
Marloi cécile	2	2	0.404898	0.595102
Flecheur florentine	2	2	0.000000	1.000000
Filou céline	2	2	0.000000	1.000000
Sié marcellin	2	2	0.387633	0.612367
Bah lise	2	2	0.000000	1.000000
Lou corentine	2	2	0.117207	0.882793
Piret émile	2	2	0.111897	0.888103
sika lazare	2	2	0.268156	0.731844
Sylla rachel	2	2	0.038128	0.961872
Piquet laurent	2	1	0.999507	0.000493
Séka clément	1	2	0.000000	1.000000
Koné abou	2	1	1.000000	0.000000
piquet sophie	1	2	0.000000	1.000000
Kassy paul	2	1	1.000000	0.000000
Baka soul	2	2	0.000000	1.000000

PARTIE II

MODELISATION

Chapitre I

MODELE DE STRUCTURATION DES DONNEES

I- Présentation du modèle

Le modèle de structuration des données sert à définir la sémantique des données qui appartiennent à la mémoire (ou base) des informations du système d'information ou qui sont véhiculées par des messages.

La structuration des informations comporte notamment la définition des données et des relations entre celles-ci.

La mémoire du système d'information comprend toutes les informations stockables sur des supports informatiques ou manuels.

Une entité est une chose concrète ou abstraite appartenant au réel perçu à propos de laquelle on veut enregistrer des informations.

Ainsi le type d'entité (T.E) est une classe de toutes les entités possibles du réel perçu qui vérifient la définition du type.

Une association est définie par une correspondance entre deux ou plusieurs entités où chacune assume un rôle donné.

Ainsi donc l'existence d'une association est contingente à l'existence des entités qu'elle met en correspondance.

Une contrainte d'intégrité est une propriété, non représentée par les concepts de base du modèle que doivent satisfaire les données appartenant à la mémoire du système d'information.

II - Description des entités et des associations

II.1- Les types d'entités.

II.1.1- Entité Patient

Définition:

Toute personne qui a subi un ou plusieurs examens cliniques pour une maladie donnée.

Attributs:

Identifiant(s):

Id1

Num_P: String; est le numéro du patient.

Id2

Nom_P: String; est le nom du patient.

Prénom_P: String; est le prénom du patient.

Date naissance_P: String; est la date de naissance du patient.

Non identifiant(s):

Sexe_P: String; est le sexe du patient. Cet attribut ne peut prendre que deux valeurs possibles:

- M (si le patient de sexe masculin) ou,
- F (si le patient de sexe féminin).

Groupe-P: String; est l'état de santé du patient. Cet attribut peut prendre deux valeurs possibles:

- 1: si le patient est sain
- 2: si le patient est malade.

Remarques

Id1: Groupe d'attribut(s) considéré(s) comme identifiant principal du type d'entité.

Id2: Groupe d'attribut(s) considéré(s) comme identifiant secondaire du type d'entité.

String: Chaîne de caractères quelconques.

II.1.2 - Entité Maladie.

Définition:

Une maladie est une altération dans la santé, dans l'équilibre des êtres humains, à propos de laquelle on décide de réaliser des tests cliniques afin de découvrir si cette maladie est susceptible d'être présente ou non chez le patient.

Les attributs:

Identifiant(s):

Id1

Nom_M: String; est le nom de la maladie.

Id2

Code_M: String; est le numéro qui permet de référencer une maladie donnée.

Non identifiant(s)

Caractéristique_M: String; est une description ou l'ensemble des symptômes qui caractérisent une maladie donnée.

II.1.3 - Entité Tests

Définition:

Un test est la réalisation d'un examen clinique dans un laboratoire médical.

Attributs:

Identifiant(s):

Id1

Nom_T: String; est le nom du test réalisé sur le patient dans un laboratoire.

II.2 - Les types d'associations.

II.2.1- Association Diagnostic clinique

Définition:

C'est la décision prise par le médecin traitant d'affecter ou non le patient à une maladie donnée sur base de tests laboratoires.

Elle associe un patient donné à une maladie concernée à propos de laquelle le patient a effectué des examens cliniques.

Attributs:

Décision: String; correspond au diagnostic posé par un clinicien qui peut ou non correspondre à celui du médecin traitant.

II.2.2 - Association Mesures.

Définition:

Associe des mesures de tests à un patient donné.

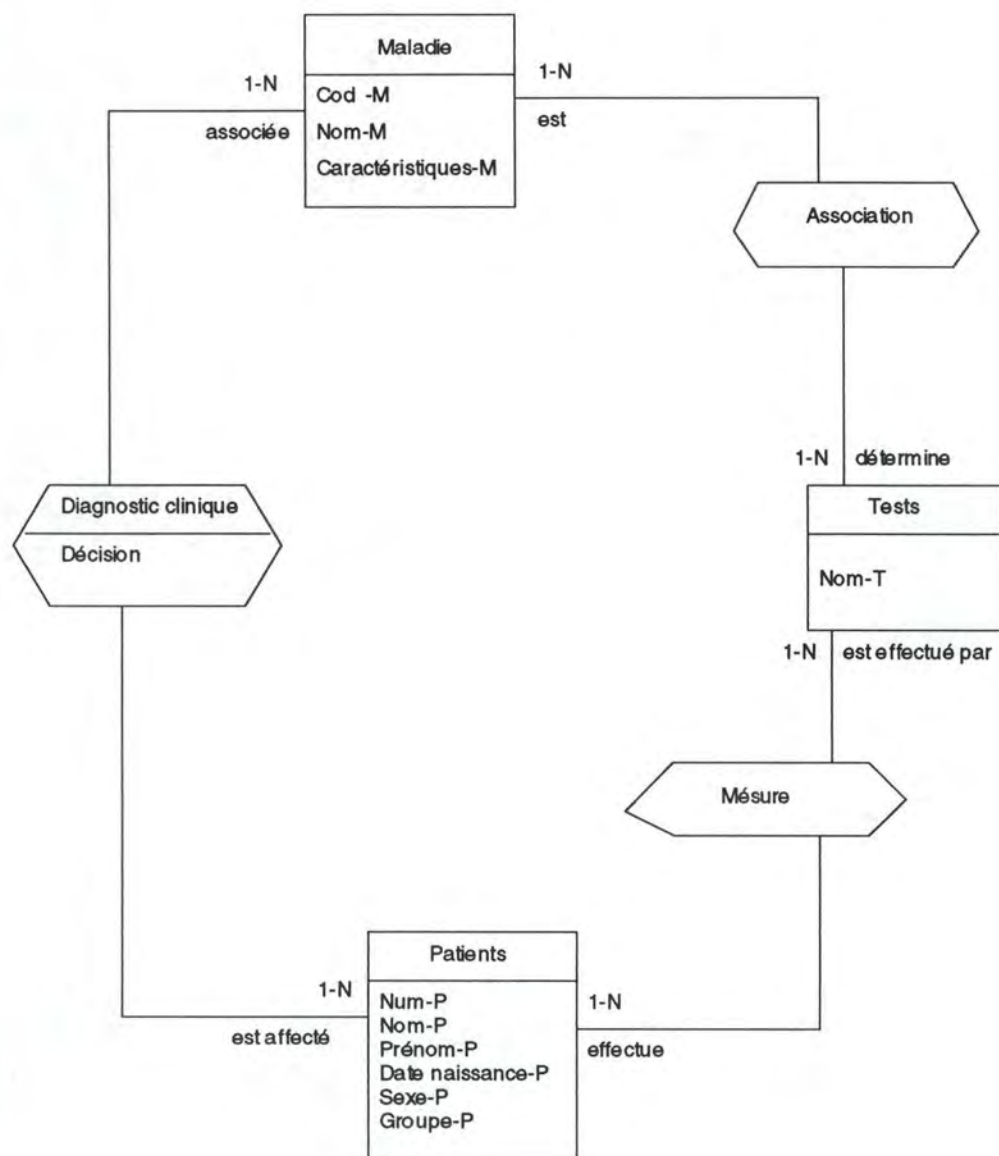
II.2.3 - Association Association.

Définition:

Associe à une maladie donnée à un ou plusieurs tests.

II.3 - Le schéma entités associations résultant.

Ce schéma a été réalisé à partir des informations ci-dessus et montre les relations entre les différents types d'entités par les types d'association définis.



III - Contraintes

III.1 - Contraintes d'intégrités fonctionnelles.

- Tout patient qui a été enregistré a subi au moins un Test.
- Tout patient qui est affecté à une maladie par le biais d'un diagnostic clinique, alors cette maladie est liée à au moins un test.

III.2 - Contraintes non fonctionnelles.

Les contraintes non fonctionnelles qui influencent certains choix relatifs à la réalisation du système sont:

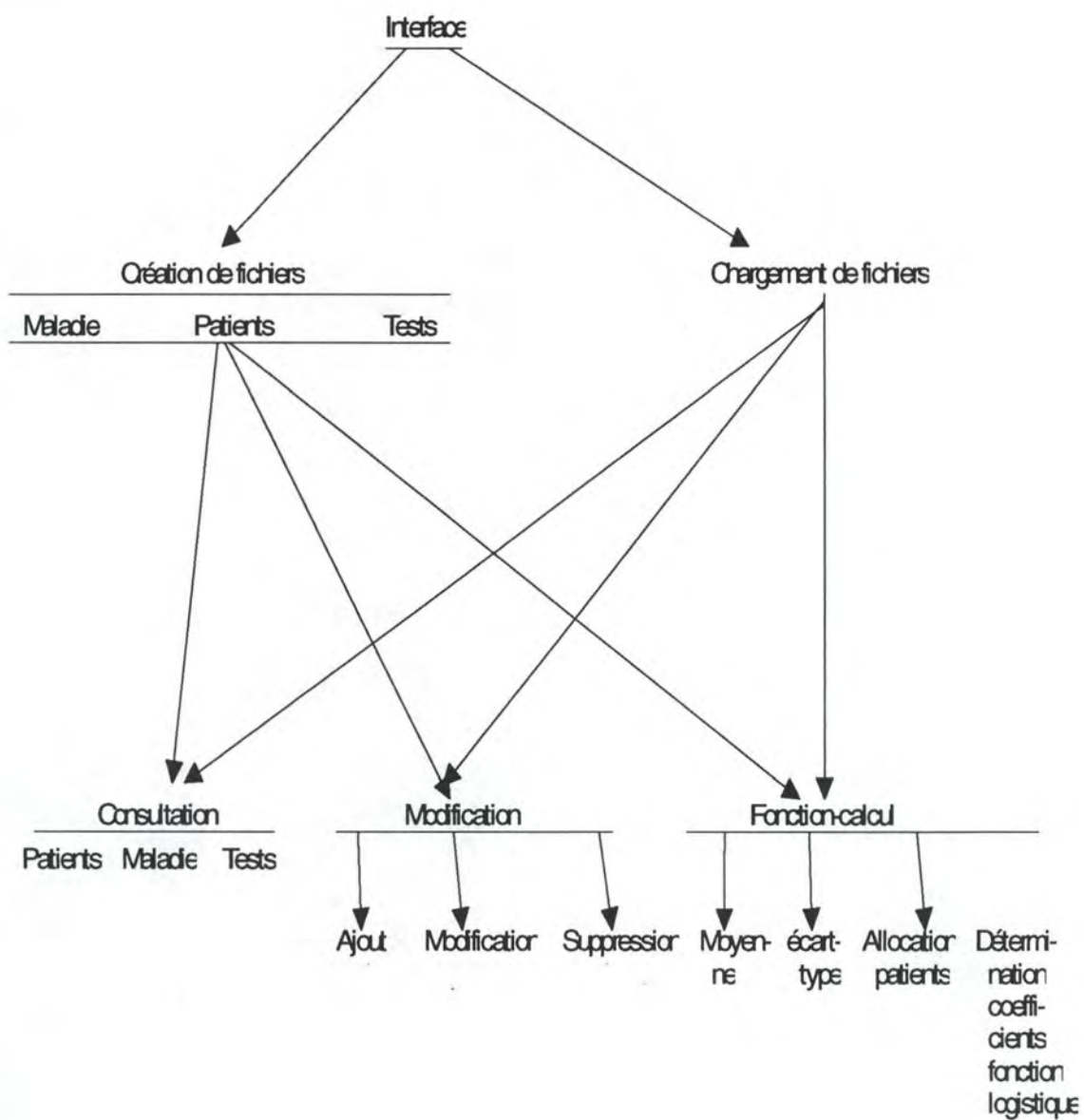
- la nécessité d'une interface conviviale,
- le système devrait être implémenté avec pascal sous Windows en utilisant la bibliothèque graphique object-windows,

IV - Scénario de fonctionnement.

Cette partie présente un exemple de fonctionnement du logiciel que nous allons concevoir. Le niveau supérieur du schéma indique l'interface c'est-à-dire la partie qui nous permet de réaliser le deuxième niveau c'est-à-dire les opérations de création et de chargement de fichiers.

Ces opérations de chargement ou de création de fichiers peuvent donner lieu à des opérations de consultation , de modification au sens large (ajout, modification ou suppression d'enregistrements) de fichiers ou de calculs.

Un exemple de scénario de fonctionnement



Chapitre II

MODELE DE STRUCTURATION DES TRAITEMENTS

I - Présentation du modèle.

Le modèle de structuration des traitements comprend un ensemble de concepts et de règles destinés à structurer un système d'information en une hiérarchie d'agrégats fonctionnels de traitements. Un agrégat comprend un ensemble de règles à suivre ou d'actions à entreprendre pour réaliser une fonctionnalité du système d'information. Chaque agrégat seront associés la définition et l'objectif de cette fonctionnalité.

II - Application à l'analyse clinique.

II.1- Niveau projet et niveau application.

L'objectif étant d'informatiser le processus d'aide au diagnostic clinique, tel qu'il est décrit dans la partie I, du document.

Ainsi l'informatisation du processus d'aide au diagnostic clinique est considéré comme projet et comme application.

II.2 - Niveau Phase et niveau fonction.

Sur base des critères fournis dans la définition, nous avons identifié les phases de l'application dégagée précédemment. Nous donnerons le nom ainsi que l'objectif de chacune d'elles.

- ***Introduction_Information:***

cette phase consiste à introduire les informations concernant un **Patient** , une **Maladie** ou un **Test**.

- ***Consultation_Information:***

cette phase vérifie l'existence d'un **Patient** , d'une **Maladie** ou d'un **Test**

- ***Annulation_Information:***

cette phase supprime les informations relatives à un **Patient**, à une **Maladie** ou à un **Test**.

- ***Changement_Information:***

cette phase consiste à modifier les informations relatives à un **Patient**, une **Maladie** ou un **Test**.

• **Fonction_calcul:**

cette phase calcule les coefficients la fonction logistique, des moyennes, de variances et s'occupe de la procedure d'allocation d'un patient dans un groupe donné etc...

Application	Fonctionnalités
Introduction_Information	Création_Maladie Création_Patient Création_Test
Consultation_Information	Consultation_Info_Maladie Consultation_Info_Patient
Annuler_Information	Suprim_Info_Maladie Suprim_Info_Patient Suprim_Info_Test
Changement_Information	Modification_Info_Patient Modification_Info_Maladie Modification_Info_Test
Fonction_Calcul	-Estimation des coefficients de la fonction vraisemblance -Détermination des moyennes et des variances -Allocation d'un patient à un groupe donné

Fonction Création

Fonction Creation_Patient.

Objectif:

introduction des informations concernant un patient lorsque les examens cliniques sont effectués sur ce dernier.

Messages d'entrée:

- Num_P : le numéro d'immatriculation du Patient concerné
- Nom_P : le nom du patient concerné
- Prenom_P: le prénom du patient
- Sexe_P: le sexe du patient
- Date_Naissance_P: la date de naissance du patient
- Groupe_P: l'état de santé du patient.
- Diagnostic_P: diagnostic du patient

Messages de sortie:

- Message: " le Patient existe déjà ".

Règle de traitement:

Il s'agit de créer une occurrence de patient à partir des informations en entrée. Dans le cas où ces informations correspondent à un patient déjà existant alors un message est alors produit.

Fonction Création_Maladie.

Objectif:

Introduction ou saisie des informations concernant une maladie

Messages entrées:

- **Code_M**: code d'identification de la maladie
- **Nom_M**: nom de la maladie
- **Caractéristiques_M**: les caractéristiques de la maladie.

Message de sortie:

- Message: Maladie existante

Action sur le S.I:

- Création d'une occurrence de maladie

Règle de traitement:

Il s'agit de créer une occurrence de maladie à partir des informations en entrée. Dans le cas où ces informations correspondent à une maladie déjà existant alors un message est alors produit.

Fonction Création_Test

Objectif:

introduction des informations concernant un Test associé à au moins une maladie prise en compte par le clinicien.

Message en entré:

- **Nom_T**: le nom du Test

Message en sortie:

- Message: "Test existe déjà".

Règles de traitement:

Il s'agit de créer une occurrence de test à partir de l'information en entrée. Dans le cas où cette information correspond à un test déjà existant alors un message est alors produit..

Fonction Suppression

Fonction Suprim_Info_Test

Objectif:

supprimer les informations relatives à un Test.

Messages en entrée:

- **Nom_T**: nom du Test

Messages en sortie:

- Message1: "Test inexistant"

Règles de traitement:

Il s'agit de supprimer une occurrence de test à partir des informations en entrée. Dans le cas où ces informations ne correspondent pas à un test déjà existant alors un message d'erreur est alors afficher.

Fonction de Suprim_Info_M.

Objectif:

supprimer les informations relatif à la maladie

Messages en entrée:

- **code_M** : code de la maladie
- ou
- **Nom_M** : nom de la maladie

Message de sortie:

- "Maladie inexistante"

Règles de traitement:

Il s'agit de supprimer une occurrence de maladie à partir des informations en entrée. Dans le cas où ces informations ne correspondent pas à une maladie déjà existant alors un message d'erreur est alors afficher.

Fonction Suprim_Info_Patient.

Objectif:

supprimer les informations relatives à un patient

Messages en entrée:

- **Num_P** : le numéro d'immatriculation du Patient concerné
- ou
- **Nom_P** : le nom du patient concerné

- **Prenom_P**: le prénom du patient
- **Date_Naissance_P**: la date de naissance du patient

Messages de sortie:

- "Patient inexistante"

Actions sur le SI:

- Supprimer une occurrence de patient.

Règles de traitement:

Il s'agit de supprimer une occurrence de patient à partir des informations en entrée. Dans le cas où ces informations ne correspondent pas à un patient déjà existant alors un message d'erreur est alors afficher.

Fonction Consultation

Fonction Consultation_Info_Patient.

Objectif:

consulter la base de données pour connaître des informations concernant un patient donné.

Messages en entrée:

- **Nom_P** : le nom du patient concerné
- **Prenom_P**: le prénom du patient
- **Date_Naissance_P**: la date de naissance du patient.

ou

- **Num_P**

Messages sortie:

- Message1: patient inexistant

ou

- Message2:
 - **Num_P** : le numéro d'immatriculation du Patient concerné
 - **Nom_P**: le nom du patient.
 - **Prenom_P**: le prénom du patient
 - **Date_Naissance_P**: la date de naissance du patient
 - **Sexe_P**: le sexe du patient
 - **Groupe_P**: l'état de santé du patient.
 - **Diagnostic_P**: diagnostic du patient

Règles de traitement:

Il s'agit de rechercher dans le système une occurrence de patient à partir des informations en entrée. Dans le cas où ces informations ne correspondent pas à un patient déjà existant alors un message d'erreur est alors afficher sinon le système fournit les informations recherchées.

Fonction **Consultation_Info_Maladie**.

Objectif:

consulter la base de données pour connaître les informations concernant une maladie donnée.

Messages en entrée:

- **Nom_M**: nom de la maladie

Message de sortie:

- "Message1": les informations suivantes sur la maladie:
 - **Nom_M**: nom de la maladie
 - **Code_M**: code d'identifiant d'une maladie
 - **Caractéristique_M**: les caractéristiques de la maladie.

ou

- "Message2": maladie inexistant

Actions sur le SI:

aucun

Règles de traitement:

Il s'agit de rechercher dans le système une occurrence de maladie à partir des informations en entrée. Dans le cas où ces informations ne correspondent pas à une maladie déjà existante alors un message d'erreur est alors afficher sinon le système fournit les informations recherchées.

Fonction Modification

Fonction **Modification_Patient**.

Objectif:

modifier les informations relatives à un patient.

Messages en entrée:

- **Nom_P** : le nom du patient concerné
- **Prenom_P**: le prénom du patient
- **Date_Naissance_P**: la date de naissance du patient

ou

- **Num_P**: le numéro du patient concerné

Messages de sortie:

Aucune

Fonction **Modification_Info_Maladie**.

Objectif:

modifier les informations relatives à une maladie

Messages en entrée:

- **Nom_M**: nom de la maladie

Messages en sortie:

- Aucune.

Fonction **Modification_Info_Test**.

Objectif:

modifier les informations relatives à un test

Messages en entrée:

- **Nom_T**: nom du test

Messages en sortie:

- Aucune.

Fonction Fonction-Calcul

Objectif:

déterminer les coefficients de la fonction logistique, les moyennes et variances, et effectuer les opérations d'affectation des patients.

Messages en entrée: rien

Messages de sortie:

- Moyennes, variance
- Coefficients de la fonction logistique
- Affectation des patients

PARTIE III

ANALYSE CONCEPTUELLE

Chapitre I

PRESENTATION GENERALE DE L'ARCHITECTURE PHYSIQUE

I - Graphe

Nous avons découpé le graphe ci-dessus de la manière suivante:

- *Au niveau 5:*

Nous disposons de trois modules. Nous avons placé tout au dessus le module Coordinateur qui contient la procédure principale du programme. Celui-ci coordonne les deux autres modules inférieurs: Fonction-calcul et Chargement.

- Le module Chargement concerne tout ce qui est chargement de fichiers.
- Le module Fonction-calcul concerne tout ce qui touche les opérations d'estimation de coefficients de la fonction logistique, la détermination de la variance, de la moyenne et l'affectation des patients à un groupe donné etc...

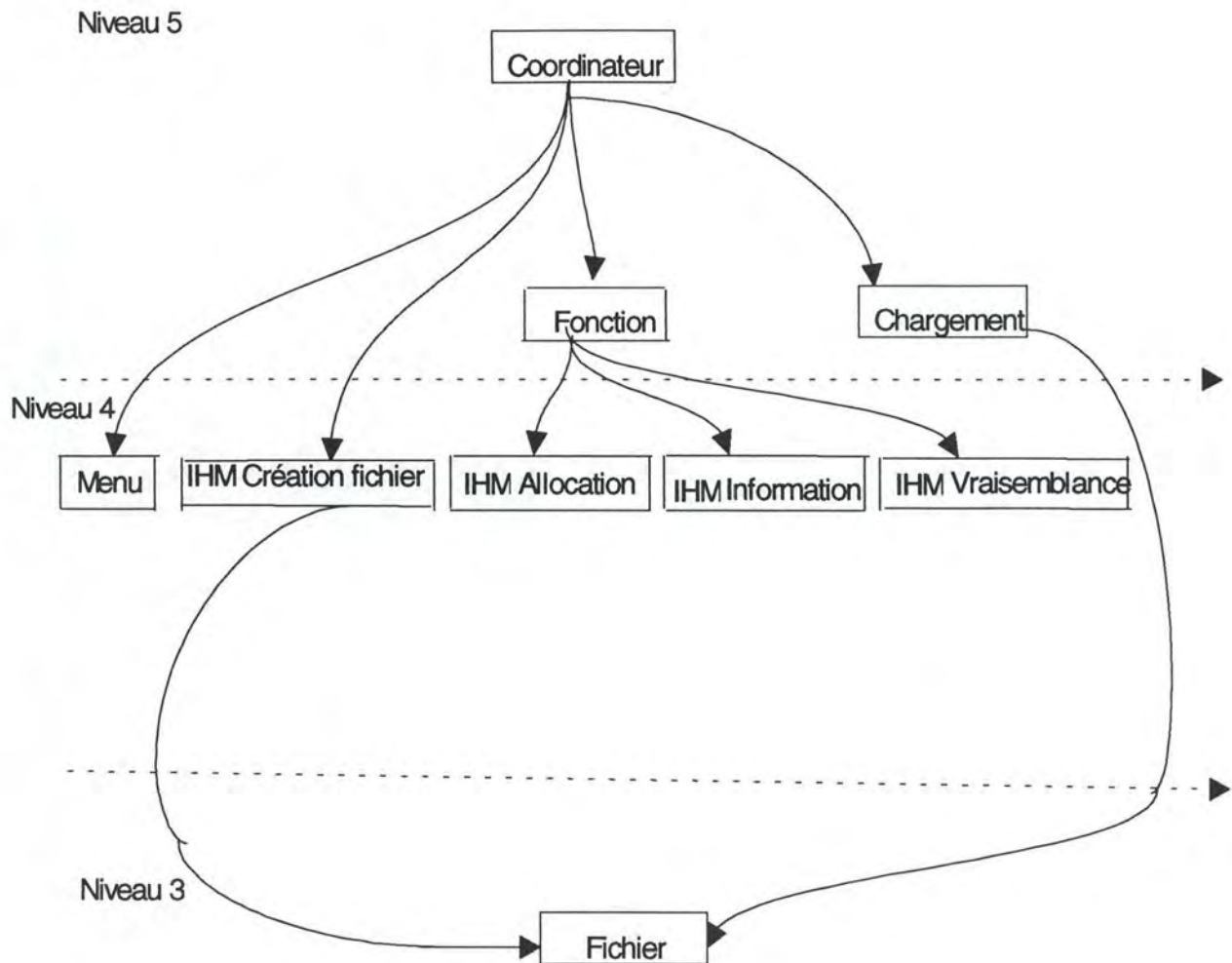
- *Au niveau 4:*

Nous disposons de cinq modules différents: Menu, IHM Création fichier, IHM Allocation, IHM Information et IHM Vraisemblance

- Le module Menu concerne l'affichage du menu principal du programme nécessaire à la gestion du logiciel
 - Le module IHM création fichier est chargé d'afficher les différents boites de dialogue nécessaire aux opérations de création de différents fichiers
 - Le module IHM Information permet l'affichage des boites dialogues nécessaires aux opérations d'ajout, de modification, de consultation ou de suppression des informations des fichiers
 - Le module IHM Allocation concerne l'affichage des boites de dialogue nécessaires opérations d'affectation des différents patients du fichier
 - Le module IHM Vraisemblance permet l'affichage des boites de dialogue nécessaires aux coefficients de la fonction de vraisemblance et à la détermination de la valeur de vraisemblance d'un individu.

- *Au niveau 3:*

Nous n'avons ici qu'un seul module c'est-à-dire le module fichier. Ce module concerne les opérations de stockage des différents fichiers sur disque.



II - Spécification algorithme détaillée

voir annexes

Chapitre II

DESCRIPTION DU FONCTIONNEMENT GENERALE DU LOGICIEL

Le lancement du logiciel entraîne inévitablement l'affichage d'un menu principal composé de cinq items de menu dont: Fichier, Informations, Vraisemblance, Allocation, et Aide. Voir schéma ci-dessous.



L'item de menu Fichier porte général sur des opérations courants de manipulations de fichiers qu'on trouve généralement sur la plupart des logiciels, comme ouvrir, fermer, quitter un fichier etc...

L'item de menu Informations porte sur des informations d'ordre générales comme le nombre total de patients, les probabilités des sains et des malades, les moyennes et ecart-types par variables tests etc....

L'item de menu Vraisemblance porte sur les coefficients de vraisemblance, la détermination de la valeur de vraisemblance d'un individu quelconque etc...

L'item de menu Allocation porte lui sur l'allocation des patients

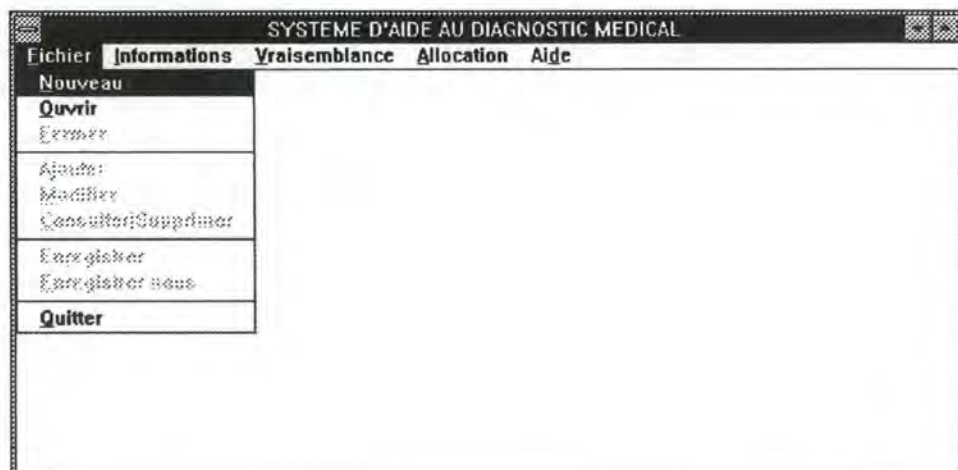
L'item de menu Aide ne spécifie que des informations d'ordre générales sur l'année de création du logiciel, le numéro de la version etc...

On accède facilement à un de ces items de menu en cliquant avec une souris sur l'item de menu choisi ou en appuyant simultanément sur les touches ALT et la lettre soulignée de l'item de menu que l'on veut utiliser(exemple si l'on veut utiliser l'item de menu fichier on utilise de façon simple, simultanément le touche ALT et la touche F)

Item menu Fichier

Le choix de cet item menu nous permet de découvrir huit items de commande (voir schéma ci-dessous)

Lorsque le logiciel est lancé tous items de commande des différents items de menu sont grisés (c'est-à-dire non accessibles) sauf les items de commande Nouveau, Ouvrir et Quitter de l'item de menu Fichier et l'item de commande A Propos de l'item de menu Aide.



Ces huit items de commande sont repartis en quatre groupes. Ces regroupements sont basés sur des critères ergonomiques dont nous ne parlerons pas ici car ce n'est pas le but de notre sujet.

Le premier groupe composé de trois items de commande, porte sur les opérations de création, d'ouverture et fermeture de fichier.

Le deuxième groupe composé de deux items de commandes, porte sur les opérations de consultation, d'ajout, de suppression et de modification des données d'un fichier.

Le troisième groupe composé lui aussi de deux items de commande, porte sur les opérations d'enregistrement de fichiers.

Le quatrième et dernier groupe est composé d'un seul item de commande qui permet de sortir définitivement du logiciel.

1) Item de commande Nouveau

Le choix cet item, entraîne irrémédiablement l'affichage de la boîte de dialogue ci-dessous



Cette boîte invite l'utilisateur à saisir une valeur entière qui représente le nombre de variables tests choisies.

Le choix du bouton de commande Cancel donne la possibilité à l'utilisateur de mettre fin à l'opération de création de fichier. Désormais, le choix d'un tel bouton donnera toujours la possibilité à l'utilisateur de mettre fin à une opération qu'il a commencé.

Le choix du bouton de commande Ok permet à l'utilisateur de continuer l'opération de création du fichier par l'affichage de la boîte de dialogue ci-dessous.

Cette boîte de dialogue porte sur la saisie des informations concernant le type d'échantillon utilisé, le nom, le code et les caractéristiques de la maladie, les noms des variables associés à cette maladie.

Nouveau Fichier

Codification

Type d'échantillon : 0 = Mélange
1 = Séparé

Type de groupe : 1 = Sain
2 = Malade

Maladie

Code :

Num :

Choix d'analyse

Type d'échantillon : ☐

Nombre de tests : ☐

Noms des variables tests

Test1: Test3: Test5:

Test2: Test4: Test6:

Caractéristiques de la maladie

OK

Cancel

Lorsque l'utilisateur clique sur le bouton Ok, cela nous conduit à la boîte de dialogue ci-dessous

Fichier

Codification
 Type de groupe : 1 Sain
 2 Malade

Patient
 N° Identif. :
 Nom :
 Prénom :
 Sexe :
 D. Naissance :

Noms des variables tests
 Test1 : Test2 :
 Test3 : Test4 :
 Test5 : Test6 :

Valeurs des variables tests
 Test1 : Test2 :
 Test3 : Test4 :
 Test5 : Test6 :

Diagnostic patient

Cette boîte de dialogue porte spécialement sur les informations concernant le patient notamment son nom, son prénom, son numéro d'identification etc...

Une fois les cases de cette boîte de dialogue remplies, lorsque l'utilisateur clique sur le bouton de commande Ajouter, cela lui donne la possibilité d'enregistrer un nouveau patient. Cette commande est utilisée lorsqu'on a beaucoup de patients à saisir.

Remarque:

- Utiliser le bouton de commande Ajouter pour saisir les informations du premier jusqu'à l'avant dernier patient. Arrivée au dernier patient à saisir utiliser le bouton de commande Ok pour progresser dans l'opération de création du nouveau fichier.
- Lorsqu'on a un seul patient à saisir seul le bouton de commande Ok est utilisé.

Lorsque le bouton de commande Ok est utilisé et que le type d'échantillon choisi par l'utilisateur est " l'échantillon mélangé " alors l'opération de création de fichier est terminée. Lorsque le bouton de commande Ok est utilisé et que le type d'échantillon choisi par l'utilisateur est "l'échantillon séparé " alors cela entraîne l'affichage de la boîte de dialogue ci-dessous. Cette boîte de dialogue porte sur les probabilités estimées du groupe sain et du groupe malade

Probabilités estimées

Sain :

Malade :

Le choix du bouton Ok après le remplissage des différentes cases termine à l'opération de création de fichier.

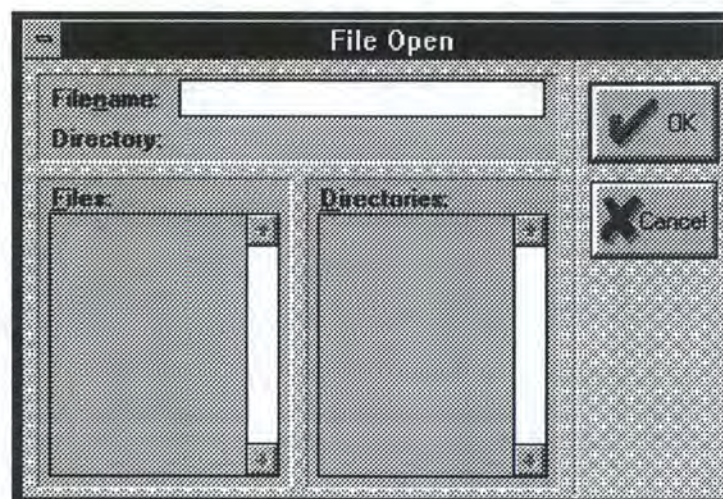
Remarque:

- Une fois l'opération de création de fichier terminée les items de commande de l'item de menu Fichier auparavant grisés sont dégrisés (accessibles) alors les items de commande Nouveau et Ouvrir de ce même item de menu fichier sont grisés (voir schéma ci-dessous).
- L'item de commande Matrice globale des données de l'item de menu Informations est aussi dégrisé
- On ne peut créer un nouveau fichier qu'en fermant le fichier déjà créé (cliquer sur l'item de commande Fermer).
- Dans la suite de l'analyse notion commande sera sous entendu item de commande et la notion de menu sera sous entendu item de menu. Ainsi en disant le menu Fichier cela doit sous entendre item de menu Fichier, de même qu'en parlant de commande Ouvrir cela doit sous entendre l'item de commande Ouvrir.



2) Item de commande Ouvrir

Cette commande favorise de chargement d'un fichier existant. Cela entraîne l'affichage de la boîte de dialogue ci-dessous



Cette boîte de dialogue affiche la liste des fichiers existants. En sélectionnant un fichier et en cliquant sur le bouton Ok, cela entraîne le chargement de ce fichier.

Lorsque le chargement est réussi le logiciel affiche une boîte de message pour informer l'utilisateur de la réussite de cette opération.

Ainsi tous les items de commande grisés sont dégrisés et les items de commande Nouveau et Ouvrir sont grisés.

Pour ouvrir un autre fichier, il faut que l'utilisateur ferme le fichier déjà ouvert.

3) Item de commande *Fermer*

Cette commande est utilisée pour fermer un fichier existant ouvert ou un fichier nouvellement créé. Lorsque cette commande est actionnée, logiciel vérifie que le fichier que l'on veut fermer est un ancien ou pas.

Si c'est un ancien fichier qui a connu des modifications le logiciel donne la possibilité à l'utilisateur de pouvoir sauvegarder son travail avant de fermer ce fichier.

Si c'est un fichier qui vient d'être créé, logiciel vérifie si ce fichier a déjà été enregistré. Si ce n'est pas le cas alors logiciel donne la possibilité à l'utilisateur de sauvegarder son travail avant la fermeture du fichier.

4) Item de commande *Ajouter*

Cette commande donne la possibilité à l'utilisateur d'augmenter la taille d'un fichier déjà existant. L'exécution de cette commande entraîne l'affichage de la boîte de dialogue ci-dessous.

On utilisera donc le bouton Ok quand on a un seul patient à ajouter au fichier ou quand on a affaire au dernier patient lorsqu'on a plusieurs patients à saisir.

Le bouton Ajouter sera utilisé si l'on a plusieurs patients à saisir.

The dialog box is titled "Fichier". It is divided into several sections:

- Codification**: Contains a label "Type de groupe :" and two radio buttons, "1 Sain" and "2 Malade".
- Patient**: Contains fields for "N° Identif.:", "Nom :", "Prénoms :", "Groupe :", "Sexe :", and "D. Naissance :".
- Noms des variables tests**: Contains six input fields labeled "Test1:", "Test2:", "Test3:", "Test4:", "Test5:", and "Test6:".
- Valeurs des variables tests**: Contains six input fields labeled "Test1:", "Test2:", "Test3:", "Test4:", "Test5:", and "Test6:".
- Diagnostic patient**: A large text area for entering the patient's diagnosis.

At the bottom of the dialog box, there are three buttons: "Ok" (with a checkmark icon), "Ajouter", and "Cancel" (with an 'X' icon).

5) Item de commande **Modifier**

Cette commande est utilisée pour modifier les informations d'un fichier existant. L'exécution de cette dite commande permet l'affichage de la boîte de dialogue ci-dessous.

Pour modifier les informations d'un patient, il suffit de saisir le numéro d'identification du patient, puis de cliquer sur le bouton Ok et toutes les informations que l'on veut modifier s'affichent dans la boîte de dialogue.

Après modification cliquer sur le bouton Modifier, qui signale au logiciel que l'on veut de tenir compte des nouvelles modifications. Ce qui a pour effet, l'effacement de toutes les boîtes de saisie, donnant ainsi la possibilité à l'utilisateur d'opérer une autre modification.

Après la dernière modification, l'utilisateur peut sortir de cette opération en cliquant sur le bouton Ok.

The screenshot shows a Windows-style dialog box titled "Fichier". It contains several input fields for patient data: "N° identifi.", "Nom", "Prénom", "Groupe", "Sexe", "D. Naissance", and a section for "Nom des variables tests" and "Valeurs des variables tests" with multiple input fields labeled Test1 through Test6. At the bottom, there is a large text area labeled "Diagnostic patient". The dialog box has three buttons at the bottom: "Ok" (with a checkmark icon), "Modifier", and "Annuler" (with an 'X' icon).

6) Item de commande **Consulter|Supprimer**

Cette commande est utilisée pour consulter ou supprimer les informations d'un fichier existant. L'exécution de cette dite commande permet l'affichage de la boîte de dialogue ci-dessous.

Pour consulter les informations d'un patient, il suffit de saisir le numéro d'identification du patient, puis de cliquer sur le bouton Consulter et toutes les informations que l'on recherche seront affichées dans la boîte de dialogue.

Pour Supprimer les informations d'un patient, il suffit de saisir le numéro d'identification du patient, puis de cliquer sur le bouton Supprimer et toutes les informations que l'on veut supprimer seront effectivement supprimées. Après modification cliquer sur le bouton Supprimer, qui signale au logiciel que l'on veut tenir compte des suppressions opérées. Ce qui a pour effet, l'effacement de la boîte de saisie du numéro d'identification, donnant ainsi la possibilité à l'utilisateur d'opérer une autre suppression. Après la consultation ou la dernière suppression, l'utilisateur peut sortir de cette opération en cliquant sur le bouton Ok.

Fichier

Patient

N° identifi. : Nom : Prénoms :

Groupe : Sexe : D. Naissance :

Noms des variables tests		Valeurs des variables tests	
Test1 : <input type="text"/>	Test2 : <input type="text"/>	Test1 : <input type="text"/>	Test2 : <input type="text"/>
Test3 : <input type="text"/>	Test4 : <input type="text"/>	Test3 : <input type="text"/>	Test4 : <input type="text"/>
Test5 : <input type="text"/>	Test6 : <input type="text"/>	Test5 : <input type="text"/>	Test6 : <input type="text"/>

Diagnostic patient

7) Item de commande **Enregistrer sous**

Cette commande permet à l'utilisateur de stocker un fichier sous un nom et sur le support de stockage de son choix.

L'exécution de cette dite commande permet l'affichage de la boîte de dialogue ci-dessous.

Lorsque l'utilisateur choisit un nom et un support de stockage, le logiciel se contente de réaliser l'opération de sauvegarde du fichier en ajoutant une extension "med" au nom de ce fichier après que l'utilisateur ait choisi le bouton Ok.

File Save As

Filename:

Directory:

Directories:

8) Item de commande *Enregistrer*

Cette commande est en générale utilisée lorsque le fichier a déjà été sauvegardé au moins une fois c'est-à-dire que ce fichier possède un nom et que l'utilisateur lui a auparavant choisi un support de stockage. Ainsi en utilisant cette commande l'utilisateur signale au logiciel qu'il veut conserver le même du fichier et le même support de stockage.

Lorsqu'un fichier nouvellement créé n'a pas été encore sauvegardé, le logiciel emmène l'utilisateur à exécuter de façon indirecte la commande Enregistrer sous en lui affichant la boîte de dialogue juste ci-dessus.

9) Item de commande *Quitter*

L'exécution de cette commande permet à l'utilisateur de sortie définitivement de l'application. Mais avant de sortie le logiciel vérifie si un fichier n'a pas été ouvert pas l'utilisateur:

- si ce n'est pas le cas le logiciel ferme alors l'application.
- si c'est le cas, le logiciel vérifie ensuite si ce fichier a subi ou pas des modifications que l'utilisateur a oublié de sauvegarder:
 - si ce n'est pas le cas, le logiciel fermer lui même le fichier et ensuite l'application.
 - si c'est le cas, le logiciel demande à l'utilisateur s'il veut sauvegarder le fichier modifié ou pas. Quel que soit la décision de l'utilisateur, il ferme lui même le fichier puis ensuite l'application

Item menu Informations

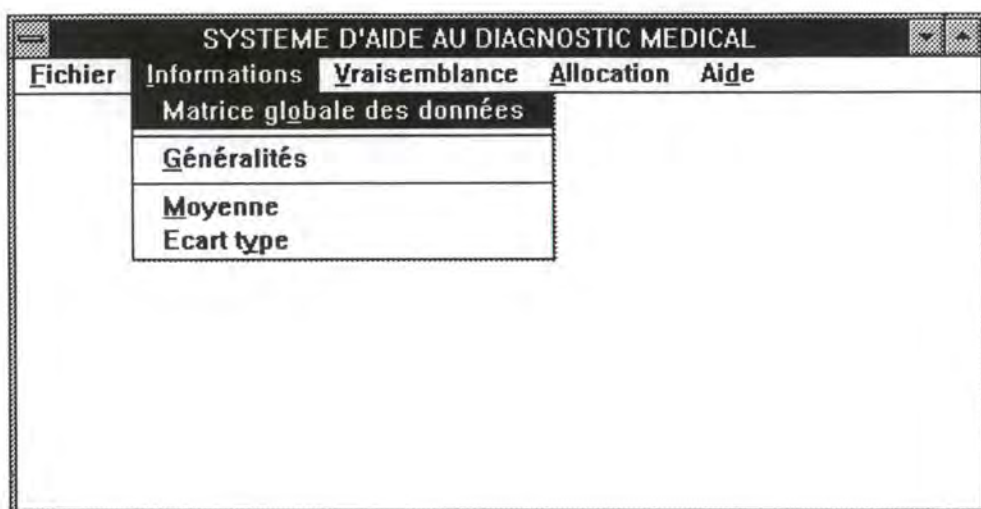
Le choix de cet item menu nous permet de découvrir quatre items de commande (voir schéma ci-dessous)

Ces quatre items de commande sont repartis en trois groupes. Ces regroupements sont aussi basées sur des critères ergonomiques.

Le premier groupe composé d'un seul item de commande, porte sur une vue globale ou synthétique de toutes les informations du fichier.

Le deuxième groupe est composé d'un seul item de commande qui permet d'afficher des informations d'ordre générales comme le nombre total de patients, le nombre de patients malades, de patients sains, les moyennes et les écart-types par variable test etc...

Le troisième et dernier groupe composé de deux items de commandes, porte sur les moyennes et les variances par type de groupe de patients et par variable test.



1) Item de commande *Matrice globale des données*

Le choix cet item, entraîne irrémédiablement l’affichage de la boîte de dialogue ci-dessous. Cette boîte permet à l’utilisateur d’avoir une vue synthétique ou globale sur l’ensembles des informations d’un fichier ouvert.

Le fait de cliquer sur le bouton Ok de cette boîte de dialogue permet de la fermer.

Identification des patients				Valeurs de tests					
Noms et Prenoms	Gr.	Sex	Date Naiss.	Test1	Test2	Test3	Test4	Test5	Test6

2) Item de commande *Généralités*

Le choix cet item, entraîne irrémédiablement l’affichage de la boîte de dialogue ci-dessous. Cette boîte permet à l’utilisateur d’avoir des informations d’ordre générales comme le nombre total de patients, le nom de la maladie, les moyennes et les ecart-types par variables tests etc...

Le fait de cliquer sur le bouton Ok de cette boîte de dialogue permet de la fermer.

Généralités

Statistiques des variables tests

Ecart type	Moyennes	Variables tests

Statistiques des patients

Nbr. de malades : Maladie :

Nbr. de sains : Proba. des malades :

Nbr. total de patients : Proba. des sains :

☒ OK

3) Item de commande *Moyenne*

Le choix cet item, entraîne irrémédiablement l'affichage de la boîte de dialogue ci-dessous. Cette boîte permet d'avoir la moyenne des variables tests par type de groupe de patients. Le fait de cliquer sur le bouton Ok de cette boîte de dialogue permet de la fermer.

Moyennes

Noms des variables tests

Test1 : Test2 : Test3 :

Test4 : Test5 : Test6 :

Variables tests

		Test 1	Test 2	Test3	Test4	Test5	Test6
Patients	Sains						
	Malades						

☒ OK

4) Item de commande *Ecart type*

Le choix cet item, entraîne irrémédiablement l'affichage de la boîte de dialogue ci-dessous. Cette boîte permet d'avoir la variance des variables tests par type de groupe de patients. Le fait de cliquer sur le bouton Ok de cette boîte de dialogue permet de la fermer.

Ecart type

Noms des variables tests

Test1 : Test3 : Test5 :

Test2 : Test4 : Test6 :

☐ OK

		Ecart type					
		Test 1	Test 2	Test3	Test4	Test5	Test6
Patients	Sains						
	Malades						

Item menu **Vraisemblance**

Le choix de cet item menu nous permet de découvrir deux items de commande (voir schéma ci-dessous).

Ces items de commandes porte sur la détermination des coefficients de vraisemblance et la détermination de la valeur de vraisemblance d'un patient quelconque donnée.

SYSTEME D'AIDE AU DIAGNOSTIC MEDICAL

Fichier Informations **Vraisemblance** Allocation Aide

Fonction de Vraisemblance

Valeur de vraisemblance

1) Item de commande *Fonction de vraisemblance*

Le choix cet item, entraîne irrémédiablement l'affichage de la boîte de dialogue ci-dessous. Cette boîte permet d'avoir les coefficients de la fonction de vraisemblance.

Le fait de cliquer sur le bouton Ok de cette boîte de dialogue permet de la fermer.

Fonction de vraisemblance

Noms des variables tests

Test1 : Test3 : Test5 :

Test2 : Test4 : Test6 :

Coefficients de la fonction de vraisemblance

Test1 : Test3 : Test5 :

Test2 : Test4 : Test6 :

Constante :

OK

2) Item de commande *Valeur de vraisemblance*

Le choix cet item, entraîne irrémédiablement l’affichage de la boîte de dialogue ci-dessous. Cette boîte permet à l’utilisateur de saisir les valeurs prises par un nouvel patient pour chacune de variables tests et dont veut déterminer le groupe d’appartenance.

Determination de la valeur de vraisemblance d'un patient

Noms des variables tests

Test1 : Test3 : Test5 :

Test2 : Test4 : Test6 :

Identification Patient

Nom : Prénom :

Valeurs des variables tests

Test1 : Test3 : Test5 :

Test2 : Test4 : Test6 :

OK Cancel

Le fait que l’utilisateur clique sur le bouton Ok après avoir rempli les cases de saisir des valeurs des variables tests et des informations sur le patient concerné, cela entraîne l’affichage de la boîte de dialogue ci-dessous.

Cette boîte de dialogue résume non seulement les informations portant sur l’identification du patient, les noms des variables tests, les valeurs des variables tests prises par ce patient et les coefficients de la fonctions de vraisemblance mais affiche aussi la valeur de vraisemblance et la probabilité à posteriori de ce patient.

Le fait de cliquer sur le bouton Ok de cette boîte de dialogue permet de la fermer.

Valeur de maximum de vraisemblance

Identification Patient

Nom: Prénom:

Noms des variables tests

Test1: Test3: Test5:
 Test2: Test4: Test6:

Valeurs des variables tests

Test1: Test3: Test5:
 Test2: Test4: Test6:

Coefficients de la fonction de vraisemblance

Test1: Test3: Test5:
 Test2: Test4: Test6:
 Constante:

Valeur du maximum de vraisemblance: Probabilité à posteriori:

Item menu Allocation

Le choix de cet item menu nous permet de découvrir deux items de commande (voir schéma ci-dessous).

Ces items de commandes porte sur l'allocation des patients de l'échantillon à un groupe donné.



1) Item de commande *Allocation abrégée*

Le choix cet item, entraîne irrémédiablement l'affichage de la boîte de dialogue ci-dessous. Cette boîte permet de voir de façon synthétique comment les patients de l'échantillon ont été alloués. Le fait de cliquer sur le bouton Ok de cette boîte de dialogue permet de la fermer

Allocation abrégée

Noms des variables tests

Test1 : Test3 : Test5 :

Test2 : Test4 : Test6 :

☒ Ok

		Patients	
		Sains	Malades
Patients	Sains		
	Malades		
	Total		

2) Item de commande *Allocation détaillée*

Le choix cet item, entraîne irrémédiablement l'affichage de la boîte de dialogue ci-dessous. Cette boîte permet de voir de façon détaillée comment les patients de l'échantillon ont été alloués. Le fait de cliquer sur le bouton Ok de cette boîte de dialogue permet de la fermer

Matrice des allocations détaillées

Estimation de probabilités

Sain :

Malade :

☒ Ok

Matrice des allocations & de probabilités					
Identification des patients		Répartit ^e . des patients		Probabilités	
N° Identif.	Noms & Prénoms	Gr. réel	Gr. alloué	Gr. sain = 1	Gr. malade = 2

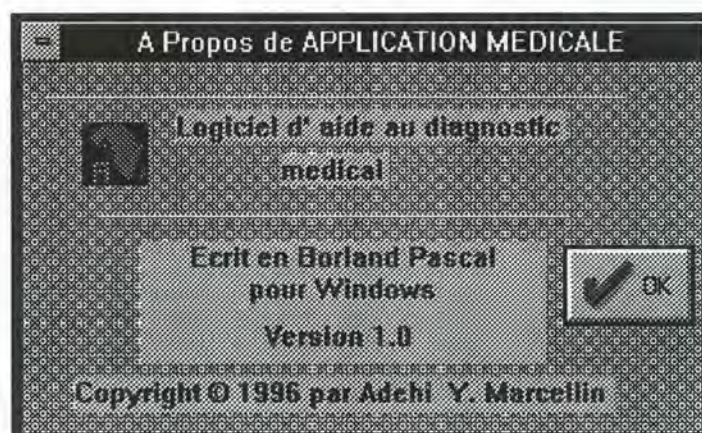
Item menu Aide

Le choix de cet item menu nous permet de découvrir un seul item de commande (voir schéma ci-dessous).



Item de commande A Propos de...

Cet item de commande porte sur les informations générales concernant l'année création du logiciel, par qui, son numéro de la version etc...



Conclusion générale:

L'utilisation de la méthode de discrimination logistique est recommandée à de nombreux égards. Tout d'abord, elle s'applique quel que soit le type de distribution envisagée, pour autant que le rapport des logarithmes des densités soit une fonction linéaire des observations. Cette condition est notamment satisfaite pour les distributions multinormales, mais aussi pour toute une série de distributions conjointes de variables discrètes et continues.

La méthode logistique conduit à une règle de classement simple.

Par définition même de la méthode, il est aisé de calculer les probabilités à posteriori.

Cette méthode est générale dans la mesure où il est applicable quel que soit le type de variables utilisées, et simple parce qu'il conduit à des fonctions discriminantes linéaires en variables.

Le logiciel conçu pour l'utilisation de cette méthode logistique discriminante a donné lieu à deux grandes étapes:

- La première étape a consisté en une étude théorique. L'objectif de cette étude étant de cerner le sujet proprement dit c'est-à-dire de comprendre en quoi consiste cette méthode
- La deuxième étape a consisté en une étape de réalisation du logiciel. Cette étape nous a conduit à suivre la méthodologie générale de conception d'un logiciel.

Le logiciel offre à l'utilisateur la possibilité de pouvoir:

- Créer plusieurs fichiers en fonction du nombre de maladies qu'il envisage étudié
- Augmenter la taille d'un fichier déjà existant
- Modifier les informations d'un fichier existant
- Réaliser les opérations d'allocation des patients, de calculer des coefficients de vraisemblance, moyennes, de variances etc...

Ce logiciel présente néanmoins quelques limites notamment:

- L'utilisateur ne peut ouvrir ou charger qu'un seul fichier à la fois. L'ouverture d'un deuxième fichier oblige ce dernier à fermer le premier fichier.
- L'utilisateur lors de la création d'un fichier ne peut dépasser plus de deux groupes et six variables tests
- L'utilisateur lors de la création d'un fichier doit obligatoirement saisir les informations sur des patients en provenance de la population des malades aussi bien que des patients en provenance de la population des sains. Si ce n'est pas le cas c'est-à-dire si l'utilisateur ne saisit que les informations uniquement que de patients malades ou bien uniquement que de patients sains le logiciel se "plante"

BIBLIOGRAPHIE

R. Couty et J.EZRA " Analyse tome 1 et 2 "

M. Queysanne " Alègre "

Gauthier-Villars Université; " Manuel de statistique biologique "

Adelin Albert et Eugène K. Harris; " Multivarite Interpretation Of Clinical Laboratory Data "

Recyclage En Statistique 4; " Analyse Discriminante "

Adelin Albert; " On the Use and Computation of Likelihood in Clinica Chemistry "

Tom Swan; " Borland Pascal 7: Programmation Windows 3.1 "

Groupe Borland; " ObjectWindows, Guide du programmeur "

ANNEXES

LES NOTIONS DE BASE DE L'ANALYSE MATHÉMATIQUE

I - Fonctions et propriétés

A) Définition:

Soient deux ensembles E et F.

On appelle f, une application sur E à valeurs dans F, qui à tout $x \in E$ associe un élément unique $y = f(x)$ dans F.

$$\begin{aligned} f: E &\rightarrow F \\ x &\rightarrow y = f(x) \end{aligned}$$

Exemple:

$$\begin{aligned} f: \mathbb{N} &\rightarrow \mathbb{R} \\ x &\rightarrow 2x + 4 \end{aligned}$$

B) Quelques propriétés:

- Soit une fonction f définie sur R, elle est dite croissante sur R si elle conserve sa relation d'ordre c'est-à-dire:

$$\forall x_1, x_2 \in \mathbb{R} \quad x_1 > x_2 \quad \Rightarrow \quad f(x_1) \geq f(x_2) \quad \text{ou} \quad \frac{f(X_1) - f(X_2)}{X_1 - X_2} \geq 0$$

Cette fonction est qualifiée de strictement croissante si:

$$\forall x_1, x_2 \in \mathbb{R} \quad x_1 > x_2 \quad \Rightarrow \quad f(x_1) > f(x_2) \quad \text{ou} \quad \frac{f(X_1) - f(X_2)}{X_1 - X_2} > 0$$

Cette fonction f sera qualifiée de décroissante si:

$$\forall x_1, x_2 \in \mathbb{R} \quad x_1 > x_2 \quad \Rightarrow \quad f(x_1) \leq f(x_2) \quad \text{ou} \quad \frac{f(X_1) - f(X_2)}{X_1 - X_2} \leq 0$$

Elle sera qualifiée de strictement décroissante si:

$$\forall x_1, x_2 \in \mathbb{R} \quad x_1 > x_2 \quad \Rightarrow \quad f(x_1) < f(x_2) \quad \text{ou} \quad \frac{f(X_1) - f(X_2)}{X_1 - X_2} < 0$$

- Toute fonction définie sur R est qualifiée de monotone sur R si elle est croissante (respectivement décroissante) sur R.

Ainsi, nous avons des fonctions monotones croissantes (respectivement décroissantes).

- Soient deux fonctions f et g définies sur R et à valeurs dans R .
On peut définir une nouvelle fonction composée de f et g .
Cette nouvelle fonction donne: $x \rightarrow g(f(x))$.
Nous noterons cette nouvelle fonction $g \circ f$.
Ainsi $(g \circ f)(x) = g(f(x))$.
Cette activité est qualifiée de composition de fonctions.

Exemple:

$$f(x) = x + 1 \text{ et } g(X) = \frac{1}{X}$$

$$(g \circ f)(x) = \frac{1}{X+1}$$

- La composition d'application est associative c'est-à-dire pour trois fonctions f , g et h définies sur R à valeurs dans R $f \circ (g \circ h) = (f \circ g) \circ h$

Cependant la composition d'applications est non commutative c'est-à-dire $f \circ g \neq g \circ f$.

- La somme de deux fonctions f et g notée $f+g$ donne

$$x \rightarrow (f+g)(x) = f(x) + g(x)$$

- Le produit d'une fonction f par un scalaire γ noté γf donne

$$x \rightarrow (\gamma f)(x) = \gamma f(x)$$

- Le produit de deux fonctions f et g noté fg donne $x \rightarrow (fg)(x) = f(x)g(x)$

- Le rapport de deux fonctions f et g noté $\frac{f}{g}$ donne

$$x \rightarrow \left(\frac{f}{g}\right)(x) = \frac{f(X)}{g(X)}$$

II - Notion de limite d'une fonction

A) Définition

Nous dirons qu'une fonction f définie au voisinage d'un point x_0 , sauf peut être en x_0 admet une limite l au point x_0 si quel que soit $\varepsilon > 0$, il existe un nombre $\alpha > 0$, tel que pour $x \neq x_0$ vérifiant $|x - x_0| < \alpha$, on ait $|f(x) - l| < \varepsilon$.
En d'autre terme $f(x)$ tend vers l quand x tend vers x_0 .

Nous écrivons alors $\lim_{x \rightarrow x_0} f(X) = l$

B) Quelques Propriétés

- $\lim (\gamma f)(x) = \gamma \cdot \lim f(x)$
- $\lim (fg)(x) = \lim f(x) \cdot \lim g(x)$
- $\lim (f + g)(x) = \lim f(x) + \lim g(x)$
- $\lim \frac{1}{f(X)} = \frac{1}{\lim f(X)}$

III - Notion de continuité de fonctions

A) Définition

On dit qu'une fonction f définie au voisinage de x_0 est dite continue en x_0 , si f admet une limite au point x_0 et que cette limite est $f(x_0)$ c'est-à-dire $\lim_{x \rightarrow x_0} f(X) = f(X_0)$

B) Quelques propriétés

- Lorsque deux fonctions f et g sont continues en un point x_0 alors $f+g$ est continue en x_0 :

$$\lim_{x \rightarrow x_0} f(X) = f(X_0) \text{ et } \lim_{x \rightarrow x_0} g(X) = g(X_0)$$

alors

$$\lim_{x \rightarrow x_0} f(X) + \lim_{x \rightarrow x_0} g(X) = f(X_0) + g(X_0)$$

alors

$$\lim_{x \rightarrow x_0} (f + g)(X) = (f + g)(X_0)$$

- Lorsque deux fonctions f et g sont continues en un point x_0 alors $f.g$ est continue en x_0 :

$$\lim_{x \rightarrow x_0} f(X) = f(X_0) \text{ et } \lim_{x \rightarrow x_0} g(X) = g(X_0)$$

alors

$$\lim_{x \rightarrow x_0} f(X) \cdot \lim_{x \rightarrow x_0} g(X) = f(X_0) \cdot g(X_0)$$

alors

$$\lim_{x \rightarrow x_0} (f \cdot g)(x) = (f \cdot g)(x_0)$$

- Lorsque deux fonctions f et g sont continues en un point x_0 alors $\frac{f}{g}$ est continue en x_0 à condition que $g(x_0) \neq 0$:

$$\lim_{x \rightarrow x_0} f(x) = f(x_0) \text{ et } \lim_{x \rightarrow x_0} g(x) = g(x_0)$$

alors

$$\frac{\lim_{x \rightarrow x_0} f(x)}{\lim_{x \rightarrow x_0} g(x)} = \frac{f(x_0)}{g(x_0)}$$

alors

$$\lim_{x \rightarrow x_0} \frac{f}{g}(x) = \frac{f}{g}(x_0)$$

- Lorsque deux fonctions f est continue en un point x_0 et g sont continues en $y = f(x_0)$ alors $f \circ g$ est continue en x_0 :

$$\lim_{x \rightarrow x_0} f(x) = f(x_0) \text{ et } \lim_{x \rightarrow x_0} g(x) = g(f(x_0))$$

alors

$$\lim_{x \rightarrow x_0} g(f(x)) = g(f(x_0))$$

alors

$$\lim_{x \rightarrow x_0} (g \circ f)(x) = (g \circ f)(x_0)$$

- Toute fonction f continue, monotone, strictement croissante ou strictement décroissante, admet une fonction réciproque noté f^{-1} tel que $f \circ f^{-1} = I$ où I est appelée fonction identité

IV - Notions de dérivées de fonctions

A) Définition

Soit une fonction f définie sur \mathbb{R} à valeur dans \mathbb{R} .

Cette fonction est dérivable en tout point x_0 si $\lim_{X \rightarrow X_0} \frac{f(X) - f(X_0)}{X - X_0}$ existe;

celle limite est notée $f'(X_0)$

En posant $h = x - x_0$ on obtient

$$\lim_{h \rightarrow 0} \frac{f(h + X_0) - f(X_0)}{h} = f'(X_0) = \frac{\delta f(h + X_0)}{\delta h}.$$

Ainsi la valeur de la nouvelle fonction au point x_0 est appelée la dérivée de la fonction f au point x_0

B) Quelques propriétés

- La dérivée la somme de deux fonctions f et g est égale à la somme des dérivées de chaque fonction:

$$(f + g)'(x) = f'(x) + g'(x)$$

- La dérivée du produit d'une fonction g par d'un scalaire λ est égale au produit de la dérivée de la fonction g par le scalaire λ :

$$(\lambda \cdot g)'(x) = \lambda \cdot g'(x) \quad \forall \lambda \in \mathbb{R};$$

- La dérivée du produit de deux fonctions f et g est égale à la somme du produit de la fonction f par la dérivée de la fonction g et du produit de la fonction g par la dérivée de la fonction f :

$$(f \cdot g)'(x) = f(x) \cdot g'(x) + f'(x) \cdot g(x);$$

- La dérivée du rapport de la fonction f par la fonction g est égale à la différence du produit de la fonction g par la dérivée de la fonction f et du produit de la fonction f par la dérivée de la fonction g , le tout divisé par le carré de la fonction g à condition que la fonction g soit non nul:

$$\left(\frac{f}{g}\right)'(X) = \frac{f'(X) \cdot g(X) - f(X) \cdot g'(X)}{(g(X))^2} \text{ avec } g(x) \neq 0;$$

- La dérivée de la composée de la fonction g à partir de la fonction f est égale à la composée de la dérivée la fonction g à partir de la fonction f le tout multiplié par la dérivée de la fonction f :

$$(g \circ f)'(x) = g' \circ f(x) \cdot f'(x)$$

V - La fonction Exponentielle et ses propriétés

A) Définition:

On appelle fonction exponentielle toute fonction f définie dans \mathbb{R} à valeur dans \mathbb{R}^+ qui à tout $x \rightarrow e^x$ ou encore notée $\exp(x)$

B) Quelques propriétés:

- La fonction exponentielle est une fonction monotone et strictement croissante sur \mathbb{R} .
- Cette fonction est aussi continue et dérivable en tout point de \mathbb{R} .

$$\lim_{x \rightarrow +\infty} \exp(x) = +\infty$$

$$\lim_{x \rightarrow -\infty} \exp(x) = 0$$

$$\exp(0) = 1$$

- Soient deux fonctions U et V définies, continues et dérivables en tout de \mathbb{R} alors la fonction f définie par la composée la fonction exponentielle par la fonction U et la fonction g définie par la composée la fonction exponentielle par la fonction V , sont toutes deux des fonctions (f et g) définies, continues et dérivables sur \mathbb{R} .

Ainsi donc:

$$\exp(U) + \exp(V) = \exp(U + V);$$

$$[\exp(U)]^n = \exp(nU) \quad \forall n \in \mathbb{N};$$

$$\frac{\delta \exp(U)}{\delta U} = \frac{\delta e^U}{\delta U}$$

$$\exp(-U) = \frac{1}{\exp(U)} \text{ avec } U \neq 0;$$

$$\exp(U - V) = \frac{\exp(U)}{\exp(V)} \text{ avec } V \neq 0;$$

$$(e^U)' = [\exp(U)]' = U' e^U = U' \exp(U)$$

$$\lim U = +\infty \Rightarrow \lim \exp(U) = +\infty$$

$$\lim U = -\infty \Rightarrow \lim \exp(U) = 0;$$

VI - La fonction logarithme népérien et ses propriétés

A) Définition:

Etant donnée que la fonction exponentielle $x \rightarrow \exp(x)$ est monotone, strictement croissante, continue et dérivable sur \mathbb{R} , alors elle admet une fonction réciproque, appelée fonction logarithme népérien et notée Log ou Ln .

Comme la fonction exponentielle est définie dans \mathbb{R} à valeur dans $\mathbb{R}^+ - \{0\}$ alors la fonction Log est définie dans $\mathbb{R}^+ - \{0\}$ à valeurs dans \mathbb{R} .

B) Quelques propriétés:

- Ainsi, tout comme la fonction $x \rightarrow \exp(x)$ la fonction Log est monotone, strictement croissante, continue et dérivable sur $\mathbb{R}^+ - \{0\}$.

$$\lim_{x \rightarrow +\infty} \text{Log}(x) = +\infty$$

$$\lim_{x \rightarrow -\infty} \text{Log}(x) = 0$$

$$\text{Log}(0) = 1$$

$$\forall x \in \mathbb{R}^* \quad \text{Log}(e^x) = x$$

$$\forall x > 0 \quad e^{\text{Log}(x)} = x$$

- Soient deux fonctions U et V définies, continues et dérivables en tout de \mathbb{R} à valeurs dans $\mathbb{R}^+ - \{0\}$ alors la fonction f définie par la composée la fonction Log par la fonction U et la fonction g définie par la composée la fonction Log par la fonction V , sont toutes deux des fonctions (f et g) définies, continues et dérivables sur \mathbb{R} .

Ainsi donc:

$$\text{Log}(U) + \text{Log}(V) = \text{Log}(U.V);$$

$$[\text{Log}(U)]^n = n.\text{Log}(U) \quad \forall n \in \mathbb{N};$$

$$\text{Log}\left(\frac{1}{U}\right) = -\text{Log}(U) \quad \text{avec } U \neq 0;$$

$$\text{Log}\left(\frac{U}{V}\right) = \text{Log}(U) - \text{Log}(V) \quad \text{avec } V \neq 0;$$

$$[\text{Log}(U)]' = \frac{U'}{U}$$

$$\lim U = +\infty \Rightarrow \lim \text{Log}(U) = +\infty$$

$$\lim U = 0 \Rightarrow \lim \text{Log}(U) = -\infty;$$

Spécification algorithme détaillée

Niveau 5

Module Coordinateur

Arrêt, FileIsChanged, LoadX: Boolean

Choix, Answer1X, id_Nouveau1X, id_Nouveau, id_Nouveau1, id_Nouveau2, decision, Button_Ok, reponse: Integer

XEchan: Char

LoadX= false

Button_Ok= 1

Tant que (not Arrêt)

case Choix

of 101: Answer1X= Application^.ExecDialog(id_Nouveau1X, TNouveau1X)

if Answer1X= Button_Ok then Answer=

Application^.ExecDialog(id_Nouveau, TNouveau)

if Answer= Button_Ok then Reponse=

Application^.ExecDialog(id_Nouveau1, TNouveau1)

if Reponse= Button_Ok and XEchan= 1 then

Decision= Application^.ExecDialog(id_Nouveau2, TNouveau2)

FileIsChanged= true

of 102: if Application^.ExecDialog(sd_FileOpen, NomFichier)= Button_Ok then

begin

LoadFile

LoadX= true

end

of 103: if not (FileIsChanged) then

begin

Answer= MessageBox('Voulez vous fermer le fichier')

if Answer=Button_Ok then fermer_fichier else garder_fichier_ouvert

end

if FileIsChanged then

begin

Answer= MessageBox('Voulez vous enregistrer le fichier avant de fermer')

if Answer= Boutton_Ok then

begin

105

fermer_fichier

end else fermer_fichier

of 104: if FileIsChanged then 105 else sauvegarder_fichier

```

of 105: if not (FileIsChanged) and LoadX then
    begin
        if Application^.ExecDialog(sd_FileSave, NomFichier)= Button_Ok then
            sauvegarder_fichier end else
            if FileIsChanged and LoadX then
                begin
                    if Application^.ExecDialog(sd_FileSave, NomFichier)= Button_Ok
                        then sauvegarder_fichier end else
                        if FileIsChanged and not (LoadX) then
                            begin
                                if Application^.ExecDialog(sd_FileSave,
                                    NouveauNomFichier)=Button_Ok
                                    then sauvegarder_fichier
                                end
                            end
                        end
                    end
                end
            end
        end
    end

of 106: Arrêt= true

of 107: Application^.ExecDialog(id_Generalite)

of 108: Application^.ExecDialog(id_Moyenne)

of 109: Application^.ExecDialog(id_Variance)

of 110: Application^.ExecDialog(id_FonctionVrai)

of 111: Answer= Application^.ExecDialog(id_ValeurVrai)
        if Answer= Button_Ok then Application^.ExecDialog(id_valeurMaxiVrai)

of 113: Application^.ExecDialog(id_AllocAbreg)

of 114: Application^.ExecDialog(id_AllocDetail)

of 115: Application^.ExecDialog(id_APropos)

of 116: Application^.ExecDialog(id_Ajouter)

of 117: Application^.ExecDialog(id_Supprimer)

of 118: Application^.ExecDialog(id_Modifier)

of 119: Application^.ExecDialog(New(id_Fichier)

```


Module Chargement

Interface

LaodFile(NomFichier): Boolean

Spécification

LaodFile(NomFichier): Boolean

MaladieX, PatientX, ProbabiliteX, TempCollection: PCollection

FileToLoad: TDosStream

FileToLaod.Init(NomFichier)

TempCollection= PCollection(FileToLaod.Get)

if TempCollection \diamond nil then MaladieX= TempCollection

TempCollection= PCollection(FileToLaod.Get)

if TempCollection \diamond nil then PatientX= TempCollection

if XEchan= 1 then

begin

TempCollection= PCollection(FileToLaod.Get)

if TempCollection \diamond nil then ProbabiliteX= TempCollection

end

LaodFile= true

Module Fonction

Interface

LogDis(MaladieX, PatientX, ProbabiliteX)

Spécification

LOGI08(Z:tabX7, NVAR:integer, var LIST:tabX2; var WT:real;

var NOBS:integer; var TOTWT:real; var XBAR, SSQ:tabX9);

si (NOBS = 0) alors

debut

TOTWT \leftarrow WT

Pour I allant de 1 à NVAR faire

debut

II \leftarrow LIST[I]

SSQ[I] \leftarrow 0

XBAR[I] \leftarrow Z[II]

fin

```

    si (WT > 0) alors NOBS ← 1
    aller au point 4
fin
SW ← TOTWT + WT
si (WT > 0) alors
    begin
        F ← WT/SW
        I ← NVAR
3: II ← LIST[I]
        D ← (Z[II] - XBAR[I])*F
        XBAR[I] ← XBAR[I] + D
        SSQ[I] ← SSQ[I] + D*TOTWT*D/F
        I ← I-1

    si (I > 0) alors aller au point 3
    TOTWT ← SW
    NOBS ← NOBS+1
4: fin

LOGI04(var A:tabX10; var M:integer; var B, C:tabX10);
    Pour I allant de 1 à M faire
        debut
            II ← (I*(I-1)) div 2
            SUM ← 0
            IK ← II
            Pour K allant de 1 à I faire
                debut
                    IK ← IK+1
                    SUM ← SUM + A[IK]*B[K]
                fin
            si (I = M) alors aller au point 50
            IK ← II + I
            I1 ← I + 1
            Pour K allant de I1 à M faire
                debut
                    IK ← IK + K - 1
                    SUM ← SUM + A[IK]*B[K]
                fin
50: C[I] ← SUM
        fin

LOGI01(var A:tabX10; var NR, NC:integer; var B:tabX10;
    M ← 1
    L20 ← 1
    Pour I allant de 1 à NCC faire

```


debut
L10 ← 1

Pour J allant de 1 à NR faire
 debut
 SUM ← 0
 L1 ← L10
 L2 ← L20
 Pour K allant 1 à NC faire
 debut
 SUM ← SUM + A[L1]*B[L2]
 L1 ← L1 + 1
 L2 ← L2 + 1
 fin
 L10 ← L10 + 1
 M ← M + 1
 fin
L20 ← L20 + NC
fin

LOGPRO(var KSI, PSI:tabX7; var KG, KG1:Integer);
CONS ← 50
KSI[KG] ← 0
Pour IS allant de 1 à KG1 faire
 debut
 PSI[IS] ← 0
 Pour IT allant de 1 à KG faire
 debut
 WORK[IT] ← KSI[IT] - KSI[IS]
 ADIF ← Abs(WORK[IT])
 si ADIF ≥ CONS alors
 debut
 SIGN ← WORK[IT]/ADIF
 WORK[IT] ← SIGN*(CONS - 1)
 fin
 fin
 fin
 Pour IT allant de 1 à KG faire PSI[IS] ← PSI[IS] + Exp(WORK[IT])
 PSI[IS] ← 1/PSI[IS]
 PSI[KG] ← 1 - PSI[IS]
 fin

```
LOGMLE(var X:tabX5; var NVAR:Integer; var N:tabX4; var KG:Integer;
        var ISDX:booleen; var ITYP:booleen;
        var PRIOR:tabX7; var CYCLE:Integer; var B:tabX10)
```

```
TANT← 0
```

```
INV←0
```

```
CIT← 0
```

```
IT← 0
```

```
KG1← KG - 1
```

```
NOBS← 0
```

```
RLLO← 0
```

```
Pour I allant de 1 à KG faire
```

```
debut
```

```
RLLO← RLLO + N[I]*Ln(N[I])
```

```
NOBS← NOBS + N[I]
```

```
fin
```

```
RLLO← RLLO - NOBS*Ln(NOBS)
```

```
NVAR1← NVAR + 1
```

```
NC←KG1*NVAR1
```

```
MM← NC*(NC + 1)div 2
```

```
L← 0
```

```
RLL← -1.E10
```

```
FINLDF← false
```

```
EPS← 1.E-6
```

```
9:
```

```
si FINLDF alors
```

```
debut
```

```
si (ITYP > 0) alors
```

```
debut
```

```
Pour T allant de 1 à KG faire
```

```
KK←1
```

```
Pour I allant de 1 à KG1 faire
```

```
debut
```

```
B[KK] ← B[KK] + Ln(N[KG]*PRIOR[I]) - Ln(N[I]*PRIOR[KG])
```

```
KK← KK + NVAR1
```

```
fin
```

```
Pour T allant de 1 à KG1 faire
```

```
debut
```



```

K ← (T-1)*NVAR1
K ← K*((K+1) div 2)
FACT ← 1
Pour T1 allant de 1 à T faire
    debut
        si (T = T1) alors FACT ← 1 + N[KG]/N[T]
        KK ← K + (T1-1)*NVAR1 + 1
        H[KK] ← H[KK] - FACT/N[KG]
    fin
fin
fin
fin

```

```

si FINLDF alors
    debut
        Pour I allant de 1 à NC faire
            debut
                II ← I
                J ← INDICE(II,II)
                SEB[I] ← SQRT(H[J])
                ZB[I] ← B[I]/SEB[I]
            fin
            CHI2 ← 2*(RLL1 - RLLO)
            IDFCHI ← KG1*NVAR
            PROCHI ← PRCHI2(CHI2,IDFCHI,IFault)
        fin
    fin

```

```

si pas(FINLDF) alors
    debut
        si CIT ≤ 25 alors
            debut
                RLL1 ← RLL
                Pour I allant de 1 à KG faire
                    Pour J allant de 1 à KG faire ERR[I,J] ← 0
                RLL ← 0
                CIT ← CIT + 1
                INVLDF ← false
                TANT ← 0
                Pour I allant de 1 à NC faire C[I] ← 0
                si (CIT-(CIT div CYCLE)*CYCLE = 1) ou (CYCLE = 1) alors
                    debut
                        INVLDF ← True
                        Pour I allant 1 à MM faire H[I] ← 0
                    fin
                K ← 1
            fin
        fin
    fin

```

```

JG ← 1
ISUM ← N[1]
Pour I1 allant de 1 à NOBS faire
  debut
    si I1 > ISUM alors
      debut
        JG ← JG + 1
        ISUM ← ISUM + N[JG]
      fin
    fin
  KK ← 1
  CHOIX ← 1
  Pour IS allant de 1 à KG1 faire
    debut
      Pour TOTO allant de 1 à NVAR1 faire
        debut
          TANT ← TANT + 1
          Z1[TOTO] ← B[TOTO]
          Q1[TOTO] ← X[TANT]
        fin
      LOGI01(Z1, CHOIX, NVAR1, Q1, CHOIX, SUMQ)
      KK ← KK + NVAR1
      KSI[IS] ← SUMQ
      SUMQ ← 0
      KSI[KG] ← 0
    fin
  LOGPRO(KSI, PSI, KG, KG1)
  RLL ← RLL + Ln(PSI[JG])
  Pour T allant de 1 à KG1 faire
    debut
      Pour J allant de 1 à NVAR1 faire
        debut
          KJ ← J
          JT ← (T - 1)*NVAR1 + J
          C[JT] ← C[JT] - PSI[T]*Q1[KJ]
          si T = JG alors C[JT] ← C[JT] + Q1[KJ]
        fin
      fin
    fin
  si INVLDF alors
    debut
      Pour T allant de 1 à KG1 faire
        debut
          Pour T1 allant de 1 à T faire
            debut
              Pour J allant de 1 à NVAR1 faire
                debut

```



```

KJ ← J
IR ← (T-1)*NVAR1 + J
JJ1 ← NVAR1
FACT ← PSI[T]*PSI[T1]*Q1[KJ]
si pas(T > T1) alors
  debut
    JJ1 ← J
    FACT ← -PSI[T]*(1-PSI[T])*Q1[KJ]
  fin

  Pour J1 allant de 1 à JJ1 faire
    debut
      KJ1 ← J1
      IR1 ← (T1 - 1)*NVAR1 + J1
      IRR1 ← INDICE(IR,IR1)
      H[IRR1] ← H[IRR1] - FACT*Q1[KJ1]
    fin
  fin
fin
fin
fin
K ← K + NVAR1
fin
si (CIT = 25) alors FINLDF ← true
LOGI04(H, NC, C, DB)
Pour I allant de 1 à NC faire B[I] ← B[I] + DB[I]
INVLDF ← false
retourner au point 9
fin
fin

```

Procedure AfficheTout1(C: PCollection;var G:integer);

```

XXNIden ← P^.Identif
XXNom ← P^.NomP
XXPrenom ← P^.PrenomP
XXGroup ← P^.Group
XXSexe ← P^.Sexe
XXDateNais ← P^.DateNais
XXTest1 ← P^.VTest1
XXTest2 ← P^.VTest2
XXTest3 ← P^.VTest3
XXTest4 ← P^.VTest
si XXGroup = IG[JK] alors
  debut
    N[JK] ← N[JK]+1
  fin

```

```

K ← K+1
Y[K] ← 1

si NV=1 alors debut
  K ← K+1
  Y[K] ← XXTest1
  L ← L+1
  NPAT[L] ← StrPas(XXNIden)
  L ← L+1
  NPAT[L] ← XXNom+ ' ' +XXPrenom
fin

si NV=2 alors debut
  K ← K+1
  Y[K] ← XXTest1
  K ← K+1
  Y[K] ← XXTest2
  L ← L+1
  NPAT[L] ← StrPas(XXNIden)
  L ← L+1
  NPAT[L] ← XXNom+ ' ' +XXPrenom
fin

si NV=3 alors debut
  K ← K+1
  Y[K] ← XXTest1
  K ← K+1
  Y[K] ← XXTest2
  K ← K+1
  Y[K] ← XXTest3
  L ← L+1
  NPAT[L] ← StrPas(XXNIden);
  L ← L+1
  NPAT[L] ← XXNom+ ' ' +XXPrenom
fin

si NV=4 alors debut
  K ← K+1
  Y[K] ← XXTest1
  K ← K+1
  Y[K] ← XXTest2
  K ← K+1
  Y[K] ← XXTest3
  K ← K+1

```



```

Y[K] ← XXTest4
L ← L+1
NPAT[L] ← StrPas(XXNIden)
L ← L+1
NPAT[L] ← XXNom+ ' ' +XXPrenom
fin

```

Procedure LogDis(MaladieX, PatientX, ProbabiliteX: PCollection)

```

CHOIX ← 1
CHOIX3 ← 6
KG ← 2
ISDX ← 0
Pour I variant de 1 à NV faire LISTX[I] ← I
NV1 ← NV + 1
L ← 0
K ← 0
KG1 ← KG - 1
CYCLE ← 1
Pour JK allant de 1 à KG faire
    debut
        LISTG[JK] ← JK
        N[JK] ← 0
        IG[JK] ← JK
        AfficheTout1(PatientX, IG[JK])
    fin;
NOBS ← 0
Pour I allant de 1 à KG faire NOBS ← NOBS + N[I]
ITYP ← XEchan
si (ITYP = 1) alors
    debut
        SUMP = 0
        SUMP ← PRIOR[1] + PRIOR[2]
    fin sinon
        debut
            Pour I allant de 1 à 2 faire
                debut
                    PRIOR[I] ← N[I]/NOBS
                    PRIOR1[I] ← PRIOR[I]
                fin
            fin
        fin
NVAR ← NV
Pour I allant de 1 à NVAR faire LIST[I] ← LISTX[I]
NTOT ← 0
K ← 0
L ← 2

```

```

MTOT←N[1]
JG← 1
JGT← 1
NT[JG] ← 0
NT1← NT[JG]
Pour I allant de 1 à NOBS faire
  debut
    KK← (I-1)*NV1 + 1
    K← K + 1
    X[K] ← Y[KK]
    Q1← NVAR11
    Pour J allant de 1 à NVAR faire
      debut
        K← K + 1
        X[K] ← Y[KK+LIST[J]]
        Z[J] ← X[K]
      fin
    fin

  %determination des moyennes XBAR et variances SD par type de test%
  LOGI08(Z,NVAR,LISTX,CHOIX,NTOT,TW,XBAR,SD)

  %determination des moyennes XBAR1 et variances SD1par type de test et type de groupe%
  LOGI08(Z,NVAR,LISTX,CHOIX,NT1,TWT1,XBAR1,SD1)
  NT[JG] ← NT1
  TWT[JG] ← TWT1
  si (I >= MTOT) alors
    debut
      JG← JG + 1
      NT← 0
      NT[JG] ← NT1
      Pour Q allant de 1 à NVAR faire
        debut
          Q1← Q1 + 1
          XB[Q1] ← XBAR1[Q]
          STD[Q1] ← SD1[Q]
        fin
      fin
      MTOT← MTOT + N[JG]
    fin
    L← L + NVAR + 1
  fin
DF← NTOT-1
Pour J allant de 1 à NVAR faire SD[J] ← Sqrt(SD[J]/DF)
Pour J allant de 1 à KG faire
  debut
    JM← (J-1)*NVAR

```



```

DF ← N[J] - 1
Pour JJ allant de 1 à NVAR faire
  debut
    JL ← JM + JJ
    STD[JL] ← Sqrt(STD[JL]/DF)
  fin
fin
NVAR1 ← NVAR + 1
NB ← KG1*NVAR1
Pour J allant de 1 à NB faire B[J] ← 0
L = 1
Pour I allant de 1 à NOBS faire
  debut
    Pour J allant de 1 à NVAR faire
      debut
        L ← L + 1
        X[L] ← (X[L] - XBAR[J])/SD[J]
      fin
    L ← L + 1
  fin
% determination des coefficients de la fonction de vraisemblance%
LOGMLE(X,NVAR,N,KG,ISDX,ITYP,PRIOR,CYCLE,B)

JG ← 1
ISUM ← N[1]
K ← 1
TANTIE ← 0
Pour I allant de 1 à NOBS faire
  debut
    si (I > ISUM) alors
      debut
        JG ← JG + 1
        ISUM ← ISUM + N[JG]
      fin
    KK ← 1
    CHOIX3 ← 1
    Pour TOTO allant de 1 à NVAR1 faire
      debut
        TANTIE ← TANTIE + 1
        Q3[TOTO] ← X[TANTIE]
      fin
    TANT ← 0
    Pour IS allant de 1 à KG1 faire
      debut
        Pour TOTO allant de 1 à NVAR1 faire
          debut

```

```

        TANT ← TANT + 1
        Z1[TOTO] ← B[TANT]
    fin
    LOGI01(Z1, CHOIX3, NVAR1, Q3, CHOIX3, SUMQ)
    KK ← KK + NVAR1
    KSI[IS] ← SUMQ
    KSI[KG] ← 0
fin
%détermination de probabilités à posteriori pour la répartition de l'échantillon par type de
groupe%
    LOGPRO(KSI, PSI, KG, KG1)
    si (JG = 1) alors
        debut
            si (PSI[1] ≥ PSI[2]) alors
                debut
                    MAX ← 1
                    ERR[JG, MAX] ← ERR[JG, MAX] + 1
                fin sinon
                    debut
                        MAX ← 2
                        ERR[JG, MAX] ← ERR[JG, MAX] + 1
                    fin
                fin
            fin
        si (JG = 2) et (PSI[1] ≤ PSI[2]) alors
            debut
                MAX ← 2
                ERR[JG, MAX] ← ERR[JG, MAX] + 1
            fin
        si (JG = 2) et (PSI[1] > PSI[2]) alors
            debut
                MAX ← 1
                ERR[JG, MAX] ← ERR[JG, MAX] + 1
            fin
        K ← K + NVAR1
    fin

```


Niveau 4

Module Menu

Interface

Affichage_Menu_Princ

Spécification

Choix = Affichage_Menu_Princ

Pas_Clique: Boolean

Choix: ENUM(101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117,

118)

- % 101 Si l'utilisateur a choisi l'item Nouveau dans le menu Fichier
- % 102 Si l'utilisateur a choisi l'item Ouvrir dans le menu Fichier
- % 103 Si l'utilisateur a choisi l'item Fermer dans le menu Fichier
- % 104 Si l'utilisateur a choisi l'item Enregistrer dans le menu Fichier
- % 105 Si l'utilisateur a choisi l'item Enregistrer-sous dans le menu Fichier
- % 106 Si l'utilisateur a choisi l'item Quitter dans le menu Fichier
- % 107 Si l'utilisateur a choisi l'item Généralités dans le menu Informations
- % 109 Si l'utilisateur a choisi l'item Moyennes dans le menu Informations
- % 110 Si l'utilisateur a choisi l'item Ecart-types dans le menu Informations
- % 111 Si l'utilisateur a choisi l'item Fonction de vraisemblance dans le menu Vraisemblance
- % 112 Si l'utilisateur a choisi l'item Valeur de vraisemblance dans le menu Vraisemblance
- % 113 Si l'utilisateur a choisi l'item Allocation abrégée dans le menu Allocation
- % 114 Si l'utilisateur a choisi l'item Allocation détaillée dans le menu Allocation
- % 115 Si l'utilisateur a choisi l'item A propos de... dans le menu Aide
- % 116 Si l'utilisateur a choisi l'item Ajouter dans le menu Fichier
- % 117 Si l'utilisateur a choisi l'item Supprimer dans le menu Fichier
- % 118 Si l'utilisateur a choisi l'item Modifier dans le menu Fichier
- % 119 Si l'utilisateur a choisi l'item Données globales dans le menu Informations

PR_MN_CREATION (100, "", "SYSTEME D'AIDE AU DIAGNOSTIC MEDICAL",
localisée, horizontale, sommet_à_la_base, textuel, fixe,
explicite, persistante, sélectionnée, ('Fichier', 'Informations',
'Vraisemblance', 'Allocation', 'Aide'), 5)

PR_MN_CREATION ('Fichier', 'F', "Fichier", localisée, verticale, sommet_à_la_base,
textuel, fixe, explicite, dynamique, glissée, ('Nouveau', 'Ouvrir',
'Fermer', 'Ajouter', 'Supprimer', 'Modifier', 'Enregistrer',
'Enregistrer-sous', 'Imprimer', 'Quitter'), 7)

PR_ITEM_AFFICHE (101, 1, textuel, 'Nouveau', 'N', fixe, actif, dialogue, false)

PR_ITEM_AFFICHE (102, 2, textuel, 'Ouvrir', 'O', fixe, actif, dialogue, false)

PR_ITEM_AFFICHE (103, 3, textuel, 'Fermer', 'F', fixe, actif, dialogue, false)

PR_ITEM_AFFICHE (-1, 4, séparateur)

PR_ITEM_AFFICHE (116, 1, textuel, 'Ajouter', 'J', fixe, actif, dialogue, false)

PR_ITEM_AFFICHE (117, 2, textuel, 'Supprimer', 'S', fixe, actif, dialogue, false)

PR_ITEM_AFFICHE (118, 3, textuel, 'Modifier', 'O', fixe, actif, dialogue, false)


```

PR_ITEM_AFFICHE (-1, 5, séparateur)
PR_ITEM_AFFICHE (104, 6, textuel, 'Enregistrer', 'R', fixe, actif, dialogue, false)
PR_ITEM_AFFICHE (105, 7, textuel, 'Enregistrer sous', 'E', fixe, actif, dialogue, false)
PR_ITEM_AFFICHE (-1, 8, séparateur)
PR_ITEM_AFFICHE (107, 9, textuel, 'Quitter', 'Q', fixe, actif, dialogue, false)

PR_MN_CREATION ('Informations', 'I', ' ', 'Information', localisée, verticale,
                sommet_à_la_base, textuel, fixe, explicite, dynamique, glissée,
                ('Données globales', 'Généralités', 'Moyennes', 'Ecart-types'), 8)
PR_ITEM_AFFICHE (119, 3, textuel, 'Données Globales', 'O', fixe, actif, dialogue, false)
PR_ITEM_AFFICHE (-1, 4, séparateur)
PR_ITEM_AFFICHE (108, 5, textuel, 'Généralités', 'G', fixe, actif, dialogue, false)
PR_ITEM_AFFICHE (-1, 6, séparateur)
PR_ITEM_AFFICHE (109, 7, textuel, 'Moyennes', 'M', fixe, actif, dialogue, false)
PR_ITEM_AFFICHE (110, 8, textuel, 'Ecart-types', 'T', fixe, actif, dialogue, false)

PR_MN_CREATION ('Vraisemblance', 'V', ' ', 'Vraisemblance', localisée, verticale,
                sommet_à_la_base, textuel, fixe, explicite, dynamique, glissée,
                ('Fonction de vraisemblance', 'Valeur de vraisemblance'), 2)
PR_ITEM_AFFICHE (111, 1, textuel, 'Fonction de vraisemblance', 'F', fixe, actif, dialogue, false)
PR_ITEM_AFFICHE (112, 2, textuel, 'Valeur de vraisemblance', 'A', fixe, actif, dialogue, false)

PR_MN_CREATION ('Allocation', 'A', ' ', 'Allocation', localisée, verticale,
                sommet_à_la_base, textuel, fixe, explicite, dynamique, glissée,
                ('Allocation abrégée', 'Allocation détaillée'), 2)
PR_ITEM_AFFICHE (113, 1, textuel, 'Allocation abrégée', 'B', fixe, actif, dialogue, false)
PR_ITEM_AFFICHE (112, 2, textuel, 'Allocation détaillée', 'D', fixe, actif, dialogue, false)

PR_MN_AFFICHAGE(100)
Pas_Clique= true

While not Pas_Clique
    if EVT_ITEM_SELECT then
        begin
            Choix= PR_MN_NUM_ITEM_SELECT
            Pas_Clique= false
        end
PR_MN_EFFACE_MENU(100)

```


Module IHM Création fichier

Interface

```
(Button_Ok, MaladieX) = TNouveau.Ok(Msg)
(Button_Ok, PatientX) = TNouveau1.Ok(Msg)
(Button_Ok, ProbabiliteX) = TNouveau2.Ok(Msg)
(Button, NbrTest) = TNouveau1X.Ok(Msg)
```

Spécification

```
(Button_Ok, NbrTest) = TNouveau1X.Ok(Msg)
PR_DBX_CREATION(1500, Modale, true, true, false, false, false, 'Fichier Nouveau', "", "")
    %Présence de deux boutons: Ok et Annuler
PR_LBL_CREATION(1503, 'Choix du nombre de tests entre 1 et 6',
    Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1502, 'Nombre de tests:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(1501, NbrTest, true, 1, 1, numérique, "", 20, "", édition, true, 1)

if EVT_DBX_BT_OK(1500)
begin
    GetText(NbrTest)
    if (NbrTest < 1) and (NbrTest < 2) and (NbrTest < 3) and (NbrTest < 4) and
        (NbrTest < 5) and (NbrTest < 6) then MessageBox('Veuillez choisir uniquement
        une valeur comprise entre 1 et 6')

    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(1500)
end

(Button_Ok, MaladieX) = TNouveau.Ok(Msg)
PR_DBX_CREATION(200, Modale, true, true, false, false, false, 'Nouveau Fichier', "", "")
    %Présence de deux boutons: Ok et Annuler
PR_BRG_CREATION('Codification', "", "", "")
PR_LBL_CREATION(214, 'Type de groupe : 1 = Sain 2 = Malade', Gauche_à_droite,
    Gauche, "", "", "")
PR_LBL_CREATION(215, 'Type d'échantillon : 0 = Melangé 1 = Séparé', Gauche_à_droite,
    Gauche, "", "", "")

PR_BRG_CREATION('Choix d'analyse', "", "", "")
PR_LBL_CREATION(216, 'Nombre de tests', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(217, 'Type d'échantillon', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(208, TEchan, true, 1, 1, numérique, "", 2, "", édition, true, 1)
PR_ED_CREATION(210, NbrTest, false, 1, 1, numérique, "", 2, "", édition, false, 1)
PR_ED_PLACE_TEXTE(210, NbrTest)

PR_BRG_CREATION('Noms des variables tests', "", "", "")
PR_LBL_CREATION(218, 'Test1:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(219, 'Test2:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(220, 'Test3:', Gauche_à_droite, Gauche, "", "", "")
```

```

PR_LBL_CREATION(221, 'Test4:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(222, 'Test5:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(223, 'Test6:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(203, NomTest1, true, 10, 10, caractère, "", 10, "", édition, false, 1)
PR_ED_CREATION(204, NomTest2, true, 10, 10, Caractère, "", 10, "", édition, false, 1)
PR_ED_CREATION(205, NomTest3, true, 10, 10, Caractère, "", 10, "", édition, false, 1)
PR_ED_CREATION(206, NomTest4, true, 10, 10, Caractère, "", 10, "", édition, false, 1)
PR_ED_CREATION(207, NomTest5, true, 10, 10, Caractère, "", 10, "", édition, false, 1)
PR_ED_CREATION(208, NomTest6, true, 10, 10, Caractère, "", 10, "", édition, false, 1)

PR_BRG_CREATION('Maladie', "", "", "")
PR_LBL_CREATION(224, 'Code:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(225, 'Nom:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(212, CodeMaladie, true, 10, 10, Alphanumérique, "", 20, "", édition, false, 1)
PR_ED_CREATION(213, NomMaladie, true, 10, 10, Caractère, "", 20, "", édition, false, 1)

```

```

if EVT_DBX_BT_OK(200) then

```

```

    begin

```

```

        GetText(NomTest1)

```

```

        GetText(NomTest2)

```

```

        GetText(NomTest3)

```

```

        GetText(NomTest4)

```

```

        GetText(NomTest5)

```

```

        GetText(NomTest6)

```

```

        GetText(TEchan)

```

```

        GetText(TAnal)

```

```

        GetText(CodeMaladie)

```

```

        GetText(NomMaladie)

```

```

        if NbrTest = 1 then

```

```

            begin

```

```

                if (NomTest2 <> '') or (NomTest3 <> '') or (NomTest4 <> '') or (NomTest5 <> '')

```

```

                or (NomTest6 <> '') then MessageBox('Veuillez respecter le nombre de tests choisis')

```

```

                    if (NomTest1='') or (TEchan='') or (TAnal='') or (CodeMaladie='') or

```

```

                    (NomMaladie='') then MessageBox('Veuillez remplir toutes les cases indiquées')

```

```

            end

```

```

        if NbrTest = 2 then

```

```

            begin

```

```

                if (NomTest3 <> '') or (NomTest4 <> '') or (NomTest5 <> '') or (NomTest6 <> '')

```

```

                then MessageBox('Veuillez respecter le nombre de tests choisis')

```

```

                if (NomTest1='') or (NomTest2) or (TEchan='') or (TAnal='') or (CodeMaladie='') or

```

```

                (NomMaladie='') then MessageBox('Veuillez remplir toutes les cases indiquées')

```

```

            end

```



```

if NbrTest = 3 then
begin
    if (NomTest4 <> ' ') or (NomTest5 <> ' ') or (NomTest6 <> ' ') then
        MsgBox('Veuillez respecter le nombre de tests choisis')
    if (NomTest1=' ') or (NomTest2) or (NomTest3) or (TEchan=' ') or (TAnal=' ')
        or (CodeMaladie=' ') or (NomMaladie=' ') then MsgBox('Veuillez remplir
                                                                    toutes les cases indiquées')

end

if NbrTest = 4 then
begin
    if (NomTest5 <> ' ') or (NomTest6 <> ' ') then MsgBox('Veuillez respecter le
                                                                    nombre de tests choisis')
    if (NomTest1=' ') or (NomTest2) or (NomTest3) or or (NomTest4) or (TEchan=' ') or
        (TAnal=' ') or (CodeMaladie=' ') or (NomMaladie=' ') then
        MsgBox('Veuillez remplir toutes les cases indiquées')
end

if NbrTest = 5 then
begin
    if (NomTest6 <> ' ') then MsgBox('Veuillez respecter le nombre de tests choisis')
    if (NomTest1=' ') or (NomTest2) or (NomTest3) or (NomTest4) or (NomTest5) or
        (TEchan=' ') or (TAnal=' ') or (CodeMaladie=' ') or (NomMaladie=' ') then
        MsgBox('Veuillez remplir toutes les cases indiquées')
end

if NbrTest = 6 then
begin
    if (NomTest1=' ') or (NomTest2) or (NomTest3) or or (NomTest4) or (NomTest5)
        or (NomTest6) or (TEchan=' ') or (TAnal=' ') or (CodeMaladie=' ') or
        (NomMaladie=' ') then MsgBox('Veuillez remplir toutes les cases indiquées')
end

MaladieX^.insert(NomTest1, NomTest2, NomTest3, NomTest4, NomTest5, NomTest6,
                  TEchan, TAnal, CodeMaladie, NomMaladie)

Button_Ok = EVT_Ok(Msg)
PR_DBX_EFFACE(200)
end

(Button_Ok, PatientX) = TNouveau1.Ok(Msg)
PR_DBX_CREATION(300, Modale, true, false, false, false, 'Nouveau Fichier', ", ", ")
    %Présence de deux boutons: Ok , Ajouter et Annuler
PR_BRG_CREATION('Codification', ", ", ")
PR_LBL_CREATION(340, 'Type de groupe : 1 = Sain 2 = Malade', Gauche_à_droite, Gauche,
                                                                    ", ", ")

```


PR_BRG_CREATION('Patient', "", "", "")
 PR_LBL_CREATION(341, 'N° identifi. :', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(342, 'Nom:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(343, 'Prénoms:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(344, 'Groupe:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(345, 'Sexe:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(346, 'D. Naissance:', Gauche_à_droite, Gauche, "", "", "")
 PR_ED_CREATION(303, NIden, true, 10, 10, Alphanumérique, "", 10, "", édition, true, 1)
 PR_ED_CREATION(304, NomPatient, true, 10, 10, Caractère, "", 20, "", édition, false, 1)
 PR_ED_CREATION(305, PrenomPatient, true, 10, 10, Caractère, "", 20, "", édition, false, 1)
 PR_ED_CREATION(308, Groupe, true, 1, 1, Caractère, "", 2, "", édition, false, 1)
 PR_ED_CREATION(307, Sexe, true, 10, 10, Caractère, "", 20, "", édition, false, 1)
 PR_ED_CREATION(306, DateNaissance, true, 8, 8, Caractère, "", 9, "", édition, false, 1)

PR_BRG_CREATION('Noms des variables tests', "", "", "")
 PR_LBL_CREATION(347, 'Test1:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(348, 'Test2:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(349, 'Test3:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(350, 'Test4:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(351, 'Test5:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(352, 'Test6:', Gauche_à_droite, Gauche, "", "", "")
 PR_ED_CREATION(315, NomTest1, false, 10, 10, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(316, NomTest2, false, 10, 10, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(317, NomTest3, false, 10, 10, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(318, NomTest4, false, 10, 10, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(319, NomTest5, false, 10, 10, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(320, NomTest6, false, 10, 10, Caractère, "", 10, "", édition, false, 1)

PR_BRG_CREATION('Valeurs des variables tests', "", "", "")
 PR_LBL_CREATION(347, 'Test1:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(348, 'Test2:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(349, 'Test3:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(350, 'Test4:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(351, 'Test5:', Gauche_à_droite, Gauche, "", "", "")
 PR_LBL_CREATION(352, 'Test6:', Gauche_à_droite, Gauche, "", "", "")
 PR_ED_CREATION(310, ValeurTest1, true, 7, 7, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(311, ValeurTest2, true, 7, 7, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(312, ValeurTest3, true, 7, 7, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(313, ValeurTest4, true, 7, 7, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(314, ValeurTest5, true, 7, 7, Caractère, "", 10, "", édition, false, 1)
 PR_ED_CREATION(321, ValeurTest6, true, 7, 7, Caractère, "", 10, "", édition, false, 1)


```

if EVT_DBX_BT_OK(300) then
begin
  GetText(NIden)
  GetText(NomPatient)
  GetText(PrenomPatient)
  GetText(DateNaissance)
  GetText(Sexe)
  GetText(Groupe)
  GetText(ValeurTest1)
  GetText(ValeurTest2)
  GetText(ValeurTest3)
  GetText(ValeurTest4)
  GetText(ValeurTest5)
  GetText(ValeurTest6)

  if NbrTest = 1 then
  begin
    if (NIden=' ') or (NomPatient) or (PrenomPatient) or (DateNaissance) or (Sexe) or
      (Groupe=' ') or (ValeurTest1=' ') then MessageBox('Veuillez remplir toutes les cases
                                                         indiquées')
  end

  if NbrTest = 2 then
  begin
    if (NIden=' ') or (NomPatient) or (PrenomPatient) or (DateNaissance) or (Sexe) or
      (Groupe) or (ValeurTest1=' ') or (ValeurTest2=' ') then
      MessageBox('Veuillez remplir toutes les cases indiquées')
  end

  if NbrTest = 3 then
  begin
    if (NIden=' ') or (NomPatient) or (PrenomPatient) or (DateNaissance) or (Sexe) or
      (Groupe) or (ValeurTest1=' ') or (ValeurTest2=' ') or (ValeurTest3=' ') then
      MessageBox('Veuillez remplir toutes les cases indiquées')
  end

  if NbrTest = 4 then
  begin
    if (NIden=' ') or (NomPatient=' ') or (PrenomPatient=' ') or (DateNaissance=' ') or
      (Sexe=' ') or (Groupe=' ') or (ValeurTest1=' ') or (ValeurTest2=' ') or
      (ValeurTest3=' ') or (ValeurTest4=' ') then
      MessageBox('Veuillez remplir toutes les cases indiquées')
  end
end

```

```

if NbrTest = 5 then
begin
    if (NIden=' ') or (NomPatient) or (PrenomPatient) or (DateNaissance) or (Sexe) or
        (Groupe) or (ValeurTest1=' ') or (ValeurTest2=' ') or (ValeurTest3=' ') or
        (ValeurTest4=' ') or (ValeurTest5=' ')
    then MessageBox('Veuillez remplir toutes les cases indiquées')
end

if NbrTest = 6 then
begin
    if (NIden=' ') or (NomPatient) or (PrenomPatient) or (DateNaissance) or (Sexe) or
        (Groupe) or (ValeurTest1=' ') or (ValeurTest2=' ') or (ValeurTest3=' ') or
        (ValeurTest4=' ') or (ValeurTest5=' ') or (ValeurTest6=' ') then
        MessageBox('Veuillez remplir toutes les cases indiquées')
end

PatientX^.insert(NIden, NomPatient, PrenomPatient, DateNaissance, Sexe, Groupe,
    ValeurTest1, ValeurTest2, ValeurTest3, ValeurTest4, ValeurTest5, ValeurTest6)

if EVT_DBX_BT_OK(300) then
begin
    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(300)
end

(Button_Ok, ProbabiliteX) = TNouveau2.Ok(Msg)
PR_DBX_CREATION(400, Modale, true, true, false, false, false, 'Probabilités estimées', ", ")
    %Présence de deux boutons: Ok et Annuler

PR_LBL_CREATION(403, 'Sain:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(404, 'Malade:', Gauche_à_droite, Gauche, ", ", ", ")
PR_ED_CREATION(401, ProbaSain, true, 10, 10, Numérique, ", 20, ", édition, true, 1)
PR_ED_CREATION(402, ProbaMalade, true, 10, 10, Numérique, ", 20, ", édition, false, 1)

if EVT_DBX_BT_OK(400) then
begin
    GetText(ProbaSain)
    GetText(ProbaMalade)
    if (ProbaSain=' ') or (ProbaMalade=' ') then MessageBox('Veuillez remplir toutes les
        cases indiquées')
    if (ProbaSain < 0) or (ProbaMalade < 0) then MessageBox('Veuillez introduire des
        nombres positifs')
    if (ProbaSain + ProbaMalade) <> 1 then MessageBox('La somme des probabilités
        doit être égale à 1')

    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(400)
end

```


Module IHM Informations

Interface

TGeneralite.SetupWindow
TMoyenne.SetupWindow
TVariance.SetupWindow

Spécification

TGeneralite.SetupWindow

PR_DBX_CREATION(500, Modale, true, true, false, false, false, 'Généralités', ", ")
 %Présence d'un bouton: Ok

PR_BRG_CREATION('Statistiques des variables tests', ", ", ")
PR_ED_CREATION(512, 'Ecart-type', false, 7, 7, Caractère, ", 10, ", édition, false, 1)
PR_MBX_CREATION(502, ", false, false, false, 800, 800, 6, 1, false, Numérique)
PR_ED_CREATION(513, 'Moyennes', false, 7, 7, Caractère, ", 10, ", édition, false, 1)
PR_MBX_CREATION(503, ", false, false, false, 800, 800, 6, 1, false, Numérique)
PR_ED_CREATION(514, 'Variables tests', false, 7, 7, Caractère, ", 10, ", édition, false, 1)
PR_MBX_CREATION(504, ", false, false, false, 800, 800, 6, 1, false, Numérique)

PR_BRG_CREATION('Statistiques des patients', ", ", ")
PR_LBL_CREATION(515, 'Nbr. de malades:', Gauche_à_droite, Gauche, ", ", ")
PR_LBL_CREATION(516, 'Nbr. de sains:', Gauche_à_droite, Gauche, ", ", ")
PR_LBL_CREATION(517, 'Nbr. total de patients:', Gauche_à_droite, Gauche, ", ", ")
PR_LBL_CREATION(518, 'Maladie:', Gauche_à_droite, Gauche, ", ", ")
PR_LBL_CREATION(519, 'Proba. des malades:', Gauche_à_droite, Gauche, ", ", ")
PR_LBL_CREATION(520, 'Proba. des sains:', Gauche_à_droite, Gauche, ", ", ")
PR_ED_CREATION(506, NbrMalades, false, 7, 7, Numérique, ", 10, ", édition, false, ")
PR_ED_CREATION(507, NbrSains, false, 7, 7, Numérique, ", 10, ", édition, false, ")
PR_ED_CREATION(508, NbrTotalPatient, false, 7, 7, Numérique, ", 10, ", édition, false, ")
PR_ED_CREATION(509, NomMaladie, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(510, ProbaMalade, false, 7, 7, Numérique, ", 10, ", édition, false, ")
PR_ED_CREATION(511, ProbaSain, false, 7, 7, Numérique, ", 10, ", édition, false, ")

While I <= NbrTest do

 begin
 PR_MBX_PLACE_TEXTE(502, Ecart-type)
 PR_MBX_PLACE_TEXTE(502, #13)
 PR_MBX_PLACE_TEXTE(502, #10)
 PR_MBX_PLACE_TEXTE(503, Moyenne)
 PR_MBX_PLACE_TEXTE(503, #13)
 PR_MBX_PLACE_TEXTE(503, #10)
 PR_MBX_PLACE_TEXTE(504, NomTest)
 PR_MBX_PLACE_TEXTE(504, #13)
 PR_MBX_PLACE_TEXTE(504, #10)
 I= I+1
 end


```

PR_ED_PLACE_TEXTE(506, NbrMalades)
PR_ED_PLACE_TEXTE(507, NbrSains)
PR_ED_PLACE_TEXTE(508, NbrTotalPatient)
PR_ED_PLACE_TEXTE(509, NomMaladie)
PR_ED_PLACE_TEXTE(510, ProbaMalade)
PR_ED_PLACE_TEXTE(511, ProbaSain)

```

```

if EVT_DBX_BT_OK(500) then
begin
    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(500)
end

```

TMoyenne.SetupWindow

```

PR_DBX_CREATION(600, Modale, true, true, false, false, false, 'Moyennes', ", ")
    %Présence d'un bouton: Ok
PR_BRG_CREATION('Noms des variables tests', ", ", ", ")
PR_LBL_CREATION(630, 'Test1:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(631, 'Test2:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(632, 'Test3:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(633, 'Test4:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(634, 'Test5:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(635, 'Test6:', Gauche_à_droite, Gauche, ", ", ", ")
PR_ED_CREATION(602, NomTest1, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(603, NomTest2, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(604, NomTest3, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(605, NomTest4, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(614, NomTest5, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(615, NomTest6, false, 7, 7, Caractère, ", 10, ", édition, false, ")

```

```

PR_ED_PLACE_TEXTE(602, NomTest1)
PR_ED_PLACE_TEXTE(603, NomTest2)
PR_ED_PLACE_TEXTE(604, NomTest3)
PR_ED_PLACE_TEXTE(605, NomTest4)
PR_ED_PLACE_TEXTE(614, NomTest5)
PR_ED_PLACE_TEXTE(615, NomTest6)

```

```

PR_MBX_CREATION(642, 'Variables tests', false, false, false, 18, 18, 2, 1, false, Caractère)
PR_ED_CREATION(636, 'Test1', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(637, 'Test2', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(638, 'Test3', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(639, 'Test4', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(640, 'Test5', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(641, 'Test6', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(643, 'Patients', false, 7, 7, Caractère, ", 30, ", édition, false, ")
PR_ED_CREATION(640, 'Sains', false, 7, 7, Caractère, ", 12, ", édition, false, ")

```



```

PR_ED_CREATION(641, 'Malades', false, 7, 7, Caractère, ", 12, ", édition, false, ")
PR_ED_CREATION(606, SainTest1, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(607, SainTest2, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(608, SainTest3, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(609, SainTest4, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(610, SainTest5, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(611, SainTest6, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(612, MaladeTest1, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(613, MaladeTest2, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(614, MaladeTest3, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(615, MaladeTest4, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(616, MaladeTest5, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(617, MaladeTest6, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_PLACE_TEXTE(606, SainTest1)
PR_ED_PLACE_TEXTE(607, SainTest2)
PR_ED_PLACE_TEXTE(608, SainTest3)
PR_ED_PLACE_TEXTE(609, SainTest4)
PR_ED_PLACE_TEXTE(610, SainTest5)
PR_ED_PLACE_TEXTE(611, SainTest6)
PR_ED_PLACE_TEXTE(612, MaladeTest1)
PR_ED_PLACE_TEXTE(613, MaladeTest2)
PR_ED_PLACE_TEXTE(614, MaladeTest3)
PR_ED_PLACE_TEXTE(615, MaladeTest4)
PR_ED_PLACE_TEXTE(616, MaladeTest5)
PR_ED_PLACE_TEXTE(617, MaladeTest6)

```

```

if EVT_DBX_BT_OK(600) then
  begin
    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(600)
  end
end

```

TVariance.SetupWindow

```

PR_DBX_CREATION(700, Modale, true, true, false, false, false, 'Ecart-types', ", ")
  %Présence d'un bouton: Ok
PR_BRG_CREATION('Noms des variables tests', ", ", ", ")
PR_LBL_CREATION(730, 'Test1:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(731, 'Test2:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(732, 'Test3:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(733, 'Test4:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(734, 'Test5:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(735, 'Test6:', Gauche_à_droite, Gauche, ", ", ", ")
PR_ED_CREATION(702, NomTest1, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(703, NomTest2, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(704, NomTest3, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(705, NomTest4, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(714, NomTest5, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(715, NomTest6, false, 7, 7, Caractère, ", 10, ", édition, false, ")

```


PR_ED_PLACE_TEXTE(702, NomTest1)
 PR_ED_PLACE_TEXTE(703, NomTest2)
 PR_ED_PLACE_TEXTE(704, NomTest3)
 PR_ED_PLACE_TEXTE(705, NomTest4)
 PR_ED_PLACE_TEXTE(714, NomTest5)
 PR_ED_PLACE_TEXTE(715, NomTest6)

PR_MBX_CREATION(742, 'Vraiables tests', false, false, false, 18, 18, 2, 1, false, Caractère)
 PR_ED_CREATION(736, 'Test1', false, 7, 7, Caractère, ", 7, ", édition, false, "
 PR_ED_CREATION(737, 'Test2', false, 7, 7, Caractère, ", 7, ", édition, false, "
 PR_ED_CREATION(738, 'Test3', false, 7, 7, Caractère, ", 7, ", édition, false, "
 PR_ED_CREATION(739, 'Test4', false, 7, 7, Caractère, ", 7, ", édition, false, "
 PR_ED_CREATION(740, 'Test5', false, 7, 7, Caractère, ", 7, ", édition, false, "
 PR_ED_CREATION(741, 'Test6', false, 7, 7, Caractère, ", 7, ", édition, false, "
 PR_ED_CREATION(743, 'Patients', false, 7, 7, Caractère, ", 30, ", édition, false, "
 PR_ED_CREATION(740, 'Sains', false, 7, 7, Caractère, ", 12, ", édition, false, "
 PR_ED_CREATION(741, 'Malades', false, 7, 7, Caractère, ", 12, ", édition, false, ")

PR_ED_CREATION(706, SainTest1, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(707, SainTest2, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(708, SainTest3, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(709, SainTest4, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(710, SainTest5, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(711, SainTest6, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(712, MaladeTest1, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(713, MaladeTest2, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(714, MaladeTest3, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(715, MaladeTest4, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(716, MaladeTest5, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_CREATION(717, MaladeTest6, false, 7, 7, Numérique, ", 7, ", édition, false, "
 PR_ED_PLACE_TEXTE(706, SainTest1)
 PR_ED_PLACE_TEXTE(707, SainTest2)
 PR_ED_PLACE_TEXTE(708, SainTest3)
 PR_ED_PLACE_TEXTE(709, SainTest4)
 PR_ED_PLACE_TEXTE(710, SainTest5)
 PR_ED_PLACE_TEXTE(711, SainTest6)
 PR_ED_PLACE_TEXTE(712, MaladeTest1)
 PR_ED_PLACE_TEXTE(713, MaladeTest2)
 PR_ED_PLACE_TEXTE(714, MaladeTest3)
 PR_ED_PLACE_TEXTE(715, MaladeTest4)
 PR_ED_PLACE_TEXTE(716, MaladeTest5)
 PR_ED_PLACE_TEXTE(717, MaladeTest6)


```

if EVT_DBX_BT_OK(700) then
begin
    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(700)
end

```

Module IHM Allocation

Interface

TAllocAbreegee.SetupWindow
TAllocDetaillee.SetupWindow

Spécification

TAllocAbrgee.SetupWindow

```

PR_DBX_CREATION(1100, Modale, true, true, false, false, false, 'Ecart-types', ", ")
    %Présence d'un bouton: Ok
PR_BRG_CREATION('Noms des variables tests', ", ", ", ")
PR_LBL_CREATION(1130, 'Test1:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(1131, 'Test2:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(1132, 'Test3:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(1133, 'Test4:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(1134, 'Test5:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(1135, 'Test6:', Gauche_à_droite, Gauche, ", ", ", ")
PR_ED_CREATION(1102, NomTest1, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(1103, NomTest2, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(1104, NomTest3, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(1105, NomTest4, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(1114, NomTest5, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(1115, NomTest6, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_PLACE_TEXTE(1102, NomTest1)
PR_ED_PLACE_TEXTE(1103, NomTest2)
PR_ED_PLACE_TEXTE(1104, NomTest3)
PR_ED_PLACE_TEXTE(1105, NomTest4)
PR_ED_PLACE_TEXTE(1114, NomTest5)
PR_ED_PLACE_TEXTE(1115, NomTest6)

PR_MBX_CREATION(742, 'Patients', false, false, false, 18, 18, 2, 1, false, Caractère)
PR_ED_CREATION(1136, 'Sains', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(1137, 'Malades', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(1138, 'Total', false, 7, 7, Caractère, ", 7, ", édition, false, ")
PR_ED_CREATION(1143, 'Patients', false, 7, 7, Caractère, ", 30, ", édition, false, ")
PR_ED_CREATION(1140, 'Sains', false, 7, 7, Caractère, ", 12, ", édition, false, ")
PR_ED_CREATION(1141, 'Malades', false, 7, 7, Caractère, ", 12, ", édition, false, ")
PR_ED_CREATION(1106, SainSains, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(1107, SainMalades, 2, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(1108, SainTotal, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(1110, MaladeSains, false, 7, 7, Numérique, ", 7, ", édition, false, ")

```



```

PR_ED_CREATION(1111, MaladeMalades,false,7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_CREATION(1112, MaladeTotal, false, 7, 7, Numérique, ", 7, ", édition, false, ")
PR_ED_PLACE_TEXTE(1106, SainSains)
PR_ED_PLACE_TEXTE(1107, SainMalades)
PR_ED_PLACE_TEXTE(1108, SainTotal)
PR_ED_PLACE_TEXTE(1110, MaladeSains)
PR_ED_PLACE_TEXTE(1111, MaladeMalades)
PR_ED_PLACE_TEXTE(1112, MaladeTotal)

```

```

if EVT_DBX_BT_OK(1100) then
  begin
    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(1100)
  end

```

TAllocDetaillee.SetupWindow

```

PR_DBX_CREATION(1200, Modale, true, true, false, false, false, 'Matrice des allocations
                                                              détaillées', ", ")

  %Présence d'un bouton: Ok
PR_BRG_CREATION('Estimation de probabilités', ", ", ")
PR_LBL_CREATION(1220, 'Sain:', Gauche_à_droite, Gauche, ", ", ")
PR_LBL_CREATION(1221, 'Malade:', Gauche_à_droite, Gauche, ", ", ")
PR_ED_CREATION(1218, ProbaSain, false, 7, 7, Numérique, ", 10, ", édition, false, ")
PR_ED_CREATION(1219, ProbaMalade, false, 7, 7, Numérique, ", 10, ", édition, false, ")
PR_ED_PLACE_TEXTE(1218, ProbaSain)
PR_ED_PLACE_TEXTE(1219, ProbaMalade)
PR_ED_CREATION(1222, 'Matrice des allocations & de probabilités', false, 30, 30,
                                                              Caractère, ", 40, ", édition, false, 1)
PR_ED_CREATION(1223, 'Identification des patients', false, 15, 15, Caractère, ", 30,
                                                              ", édition, false, 1)
PR_ED_CREATION(1224, 'Répartit° des patients', false, 13, 13, Caractère, ", 20, ",
                                                              édition, false, 1)
PR_ED_CREATION(1225, 'Probabilités', false, 15, 15, Caractère, ", 30, ", édition, false, 1)
PR_ED_CREATION(1226, 'N° Identif', false, 15, 15, Caractère, ", 20, ", édition, false, 1)
PR_ED_CREATION(1227, 'Noms & Prenoms', false, 15, 15, Caractère, ", 20, ", édition, false, 1)
PR_ED_CREATION(1228, 'Gr. réel', false, 15, 15, Caractère, ", 20, ", édition, false, 1)
PR_ED_CREATION(1229, 'Gr. alloué', false, 15, 15, Caractère, ", 20, ", édition, false, 1)
PR_ED_CREATION(1230, 'Gr. sain = 1', false, 15, 15, Caractère, ", 20, ", édition, false, 1)
PR_ED_CREATION(1231, 'Gr. malade = 2', false, 15, 15, Caractère, ", 20, ", édition, false, 1)
PR_MBX_CREATION(1211, ", false, false, false, 800, 800, 6, 1, false, Alphanumérique)
PR_MBX_CREATION(1212, ", false, false, false, 800, 800, 6, 1, false, Caratère)
PR_MBX_CREATION(1213, ", false, false, false, 800, 800, 6, 1, false, Alphanumérique)
PR_MBX_CREATION(1214, ", false, false, false, 800, 800, 6, 1, false, Alphanumérique)
PR_MBX_CREATION(1215, ", false, false, false, 800, 800, 6, 1, false, Numérique)
PR_MBX_CREATION(1216, ", false, false, false, 800, 800, 6, 1, false, Numérique)
PR_BDF_CREATION(1217, 1, 800, 1, verticale, base, 1, 6)

```



```

While I <= NbrTotalPatient do
  begin
    PR_MBX_PLACE_TEXTE(1211, NIden)
    PR_MBX_PLACE_TEXTE(1211, #13)
    PR_MBX_PLACE_TEXTE(1211, #10)
    PR_MBX_PLACE_TEXTE(1212, NomPatient + PrenomPatient)
    PR_MBX_PLACE_TEXTE(1212, #13)
    PR_MBX_PLACE_TEXTE(1212, #10)
    PR_MBX_PLACE_TEXTE(1213, Groupe)
    PR_MBX_PLACE_TEXTE(1213, #13)
    PR_MBX_PLACE_TEXTE(1213, #10)
    PR_MBX_PLACE_TEXTE(1214, GroupeAlloue)
    PR_MBX_PLACE_TEXTE(1214, #13)
    PR_MBX_PLACE_TEXTE(1214, #10)
    PR_MBX_PLACE_TEXTE(1215, ProbabiliteSain)
    PR_MBX_PLACE_TEXTE(1215, #13)
    PR_MBX_PLACE_TEXTE(1215, #10)
    PR_MBX_PLACE_TEXTE(1216, ProbabiliteMalade)
    PR_MBX_PLACE_TEXTE(1216, #13)
    PR_MBX_PLACE_TEXTE(1216, #10)
    I= I+1
  end

if EVT_DBX_BT_OK(1200) then
  begin
    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(1200)
  end

```

Module IHM Vraisemblance

Interface

```

TFonctionMaxiVraisemblance.SetupWindow
TDeterminationVraisemblance.SetupWindow
TValeurMaxiVraisemblance.SetupWindow

```

Spécification

```

TFonctionMaxiVraisemblance.SetupWindow
PR_DBX_CREATION(800, Modale, true, true, false, false, false, 'Fonction de Vraisemblance',
", ")

```

%Présence d'un bouton: Ok

```
PR_BRG_CREATION('Noms des variables tests', "", "", "")
PR_LBL_CREATION(820, 'Test1:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(821, 'Test2:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(822, 'Test3:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(823, 'Test4:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(824, 'Test5:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(825, 'Test6:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(802, NomTest1, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(803, NomTest2, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(804, NomTest3, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(805, NomTest4, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(814, NomTest5, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(815, NomTest6, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_PLACE_TEXTE(802, NomTest1)
PR_ED_PLACE_TEXTE(803, NomTest2)
PR_ED_PLACE_TEXTE(804, NomTest3)
PR_ED_PLACE_TEXTE(805, NomTest4)
PR_ED_PLACE_TEXTE(814, NomTest5)
PR_ED_PLACE_TEXTE(815, NomTest6)

PR_BRG_CREATION('Coefficients de la fonction de vraisemblance', "", "", "")
PR_LBL_CREATION(826, 'Test1:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(827, 'Test2:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(828, 'Test3:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(829, 'Test4:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(830, 'Test5:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(831, 'Test6:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(832, 'Constante:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(807, CoeficienTest1, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(808, CoeficienTest2, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(809, CoeficienTest3, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(810, CoeficienTest4, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(812, CoeficienTest5, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(813, CoeficienTest6, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(811, CoeficienConstant, false, 7, 7, Caractère, "", 10, "", édition, false, "")

PR_ED_PLACE_TEXTE(807, CoeficienTest1)
PR_ED_PLACE_TEXTE(808, CoeficienTest2)
PR_ED_PLACE_TEXTE(809, CoeficienTest3)
PR_ED_PLACE_TEXTE(810, CoeficienTest4)
PR_ED_PLACE_TEXTE(812, CoeficienTest5)
PR_ED_PLACE_TEXTE(813, CoeficienTest6)
PR_ED_PLACE_TEXTE(811, CoeficienConstant)
```



```

if EVT_DBX_BT_OK(800) then
begin
    Button_Ok = EVT_Ok(Msg)
    PR_DBX_EFFACE(800)
end

```

TDeterminationVraisemblance.SetupWindow

```

PR_DBX_CREATION(900, Modale, true, true, false, false, false, 'Determination de la
                                                                    valeur de vraisemblance d'un patient', ", ")

```

 %Présence d'un bouton: Ok et Annuler

```

PR_BRG_CREATION('Noms des variables tests', ", ", ", ")
PR_LBL_CREATION(920, 'Test1:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(921, 'Test2:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(922, 'Test3:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(923, 'Test4:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(924, 'Test5:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(925, 'Test6:', Gauche_à_droite, Gauche, ", ", ", ")

```

```

PR_ED_CREATION(902, NomTest1, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(903, NomTest2, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(904, NomTest3, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(905, NomTest4, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(914, NomTest5, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(915, NomTest6, false, 7, 7, Caractère, ", 10, ", édition, false, ")

```

```

PR_ED_PLACE_TEXTE(902, NomTest1)
PR_ED_PLACE_TEXTE(903, NomTest2)
PR_ED_PLACE_TEXTE(904, NomTest3)
PR_ED_PLACE_TEXTE(905, NomTest4)
PR_ED_PLACE_TEXTE(914, NomTest5)
PR_ED_PLACE_TEXTE(915, NomTest6)

```

```

PR_BRG_CREATION('Valeurs des variables tests', ", ", ", ")
PR_LBL_CREATION(926, 'Test1:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(927, 'Test2:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(928, 'Test3:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(929, 'Test4:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(930, 'Test5:', Gauche_à_droite, Gauche, ", ", ", ")
PR_LBL_CREATION(931, 'Test6:', Gauche_à_droite, Gauche, ", ", ", ")
PR_ED_CREATION(907, ValeurTest1, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(908, ValeurTest2, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(909, ValeurTest3, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(910, ValeurTest4, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(911, ValeurTest5, false, 7, 7, Caractère, ", 10, ", édition, false, ")
PR_ED_CREATION(916, ValeurTest6, false, 7, 7, Caractère, ", 10, ", édition, false, ")

```

```

PR_BRG_CREATION('Identification Patient', "", "", "")
PR_LBL_CREATION(932, 'Nom:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(933, 'Prénoms:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(912, NomPatient, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(913, PrenomPatient, false, 7, 7, Caractère, "", 10, "", édition, false, "")

```

```

if EVT_DBX_BT_OK(900) then

```

```

    begin
        GetText(ValeurTest1)
        GetText(ValeurTest2)
        GetText(ValeurTest3)
        GetText(ValeurTest4)
        GetText(ValeurTest5)
        GetText(ValeurTest6)
        GetText(NomPatient)
        GetText(Prenoms)
        Button_Ok = EVT_Ok(Msg)
        PR_DBX_EFFACE(900)
    end

```

```

TValeurMaxiVraisemblance.SetupWindow

```

```

    PR_DBX_CREATION(1000, Modale, true, true, false, false, false, 'Valeur de
                                                                    maximum de vraisemblance', "", "")

```

```

    %Présence d'un bouton: Ok

```

```

    PR_BRG_CREATION('Identification Patient', "", "", "")
    PR_LBL_CREATION(1030, 'Nom:', Gauche_à_droite, Gauche, "", "", "")
    PR_LBL_CREATION(1031, 'Prénoms:', Gauche_à_droite, Gauche, "", "", "")
    PR_ED_CREATION(1019, NomPatient, false, 7, 7, Caractère, "", 10, "", édition, false, "")
    PR_ED_CREATION(1020, PrenomPatient, false, 7, 7, Caractère, "", 10, "", édition, false, "")
    PR_ED_PLACE_TEXTE(1019, NomPatient)
    PR_ED_PLACE_TEXTE(1020, PrenomPatient)

```

```

PR_BRG_CREATION('Noms des variables tests', "", "", "")
PR_LBL_CREATION(1032, 'Test1:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1033, 'Test2:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1034, 'Test3:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1035, 'Test4:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1036, 'Test5:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1037, 'Test6:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(1002, NomTest1, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1003, NomTest2, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1004, NomTest3, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1005, NomTest4, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1021, NomTest5, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1022, NomTest6, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_PLACE_TEXTE(1002, NomTest1)
PR_ED_PLACE_TEXTE(1003, NomTest2)
PR_ED_PLACE_TEXTE(1004, NomTest3)

```


PR_ED_PLACE_TEXTE(1005, NomTest4)
PR_ED_PLACE_TEXTE(1021, NomTest5)
PR_ED_PLACE_TEXTE(1022, NomTest6)

PR_BRG_CREATION('Valeurs des variables tests', "", "", "")
PR_LBL_CREATION(1038, 'Test1:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1039, 'Test2:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1040, 'Test3:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1041, 'Test4:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1042, 'Test5:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1043, 'Test6:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(1007, ValeurTest1, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1008, ValeurTest2, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1009, ValeurTest3, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1010, ValeurTest4, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1019, ValeurTest5, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1020, ValeurTest6, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_PLACE_TEXTE(1007, ValeurTest1)
PR_ED_PLACE_TEXTE(1008, ValeurTest2)
PR_ED_PLACE_TEXTE(1009, ValeurTest3)
PR_ED_PLACE_TEXTE(1010, ValeurTest4)
PR_ED_PLACE_TEXTE(1019, ValeurTest5)
PR_ED_PLACE_TEXTE(1020, ValeurTest6)

PR_BRG_CREATION('Coefficients de la fonction de vraisemblance', "", "", "")
PR_LBL_CREATION(1044, 'Test1:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1045, 'Test2:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1046, 'Test3:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1047, 'Test4:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1048, 'Test5:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1049, 'Test6:', Gauche_à_droite, Gauche, "", "", "")
PR_LBL_CREATION(1050, 'Constante:', Gauche_à_droite, Gauche, "", "", "")
PR_ED_CREATION(1012, CoeficienTest1, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1013, CoeficienTest2, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1014, CoeficienTest3, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1015, CoeficienTest4, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1016, CoeficienTest5, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1017, CoeficienTest6, false, 7, 7, Caractère, "", 10, "", édition, false, "")
PR_ED_CREATION(1018, CoeficienConstant, false, 7, 7, Caractère, "", 10, "", édition, false, "")

PR_ED_PLACE_TEXTE(1012, CoeficienTest1)
PR_ED_PLACE_TEXTE(1013, CoeficienTest2)
PR_ED_PLACE_TEXTE(1014, CoeficienTest3)
PR_ED_PLACE_TEXTE(1015, CoeficienTest4)
PR_ED_PLACE_TEXTE(1016, CoeficienTest5)
PR_ED_PLACE_TEXTE(1017, CoeficienTest6)
PR_ED_PLACE_TEXTE(1018, CoeficienConstant)

Niveau 3

Module Fichier

Interface

TMaladie.Store(S: TStream)
TPatient.Store(S: TStream)
TProbabilite.Store(S: Tstream)

Spécification

TMaladie.Store(S: TStream)
 S.Write(NomTest1,SizeOf(NomTest1))
 S.Write(NomTest2,SizeOf(NomTest2))
 S.Write(NomTest3,SizeOf(NomTest3))
 S.Write(NomTest4,SizeOf(NomTest4))
 S.Write(NomTest5,SizeOf(NomTest5))
 S.Write(NomTest6,SizeOf(NomTest6))
 S.Write(Echan,SizeOf(Echan))
 S.Write(Anal,SizeOf(Anal))
 S.Write(NbrTest,SizeOf(NbrTest))
 S.Write(Code,SizeOf(Code))
 S.Write(Nom,SizeOf(Nom))

TPatient.Store(S: TStream)
 S.Write(Identif,SizeOf(Identif))
 S.Write(NomPatient,SizeOf(NomPatient))
 S.Write(PrenomPatient,SizeOf(PrenomPatient))
 S.Write(Groupe,SizeOf(Groupe))
 S.Write(Sexe,SizeOf(Sexe))
 S.Write(DateNaissance,SizeOf(DateNaissance))
 S.Write(ValeurTest1,SizeOf(ValeurTest1))
 S.Write(ValeurTest2,SizeOf(ValeurTest2))
 S.Write(ValeurTest3,SizeOf(ValeurTest3))
 S.Write(ValeurTest4,SizeOf(ValeurTest4))
 S.Write(NomTest5,SizeOf(NomTest5))
 S.Write(NomTest6,SizeOf(NomTest6))

TProbabilite.Store(S: TStream)
 S.Write(ProbaSain,SizeOf(ProbaSain))
 S.Write(ProbaMalade,SizeOf(ProbaMalade))